

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :	A2	(11) International Publication Number: WO 98/54963 (43) International Publication Date: 10 December 1998 (10.12.98)
A01N 37/18, 43/04, C12Q 1/00, 1/02, 1/68, C12N 5/00, 5/06, 15/00, 15/06, 15/09, 15/10, 15/11, G01N 33/53		
(21) International Application Number: PCT/US98/11422		Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). FLORENCE, Charles [US/US]; (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FAN, Ping [CN/US]; Apartment 302, 335 West Side Drive, Gaithersburg, MD 20878 (US). WEI, Ying-Fei [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US).
(22) International Filing Date: 4 June 1998 (04.06.98)		
(30) Priority Data: 60/048,915 6 June 1997 (06.06.97) US 60/048,882 6 June 1997 (06.06.97) US (Continued on the following page)		
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).		
(72) Inventors; and		
(75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316,		(74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US).
		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
		Published <i>With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.</i>
(54) Title: 207 HUMAN SECRETED PROTEINS		

(Continued)

60/048,892	6 June 1997 (06.06.97)	US	60/057,651	5 September 1997 (05.09.97)	US
60/048,901	6 June 1997 (06.06.97)	US	60/057,769	5 September 1997 (05.09.97)	US
60/048,900	6 June 1997 (06.06.97)	US	60/057,643	5 September 1997 (05.09.97)	US
60/048,893	6 June 1997 (06.06.97)	US	60/057,645	5 September 1997 (05.09.97)	US
60/048,964	6 June 1997 (06.06.97)	US	60/057,668	5 September 1997 (05.09.97)	US
60/048,884	6 June 1997 (06.06.97)	US	60/057,635	5 September 1997 (05.09.97)	US
60/048,894	6 June 1997 (06.06.97)	US	60/057,627	5 September 1997 (05.09.97)	US
60/048,971	6 June 1997 (06.06.97)	US	60/057,667	5 September 1997 (05.09.97)	US
60/048,885	6 June 1997 (06.06.97)	US	60/057,666	5 September 1997 (05.09.97)	US
60/049,375	6 June 1997 (06.06.97)	US	60/057,764	5 September 1997 (05.09.97)	US
60/048,881	6 June 1997 (06.06.97)	US	60/057,644	5 September 1997 (05.09.97)	US
60/048,880	6 June 1997 (06.06.97)	US	60/057,765	5 September 1997 (05.09.97)	US
60/048,896	6 June 1997 (06.06.97)	US	60/057,762	5 September 1997 (05.09.97)	US
60/049,020	6 June 1997 (06.06.97)	US	60/057,775	5 September 1997 (05.09.97)	US
60/048,876	6 June 1997 (06.06.97)	US	60/057,634	5 September 1997 (05.09.97)	US
60/048,895	6 June 1997 (06.06.97)	US	60/057,777	5 September 1997 (05.09.97)	US
60/049,019	6 June 1997 (06.06.97)	US	60/057,628	5 September 1997 (05.09.97)	US
60/048,916	6 June 1997 (06.06.97)	US	60/057,776	5 September 1997 (05.09.97)	US
60/048,970	6 June 1997 (06.06.97)	US	60/057,760	5 September 1997 (05.09.97)	US
60/048,972	6 June 1997 (06.06.97)	US	60/057,761	5 September 1997 (05.09.97)	US
60/048,949	6 June 1997 (06.06.97)	US	60/057,771	5 September 1997 (05.09.97)	US
60/048,974	6 June 1997 (06.06.97)	US	60/057,770	5 September 1997 (05.09.97)	US
60/048,883	6 June 1997 (06.06.97)	US	60/057,649	5 September 1997 (05.09.97)	US
60/048,897	6 June 1997 (06.06.97)	US	60/057,774	5 September 1997 (05.09.97)	US
60/048,898	6 June 1997 (06.06.97)	US	60/057,648	5 September 1997 (05.09.97)	US
60/049,373	6 June 1997 (06.06.97)	US	60/057,642	5 September 1997 (05.09.97)	US
60/048,917	6 June 1997 (06.06.97)	US	60/057,629	5 September 1997 (05.09.97)	US
60/048,962	6 June 1997 (06.06.97)	US	60/057,778	5 September 1997 (05.09.97)	US
60/048,878	6 June 1997 (06.06.97)	US	60/057,763	5 September 1997 (05.09.97)	US
60/049,374	6 June 1997 (06.06.97)	US	60/057,584	5 September 1997 (05.09.97)	US
60/048,875	6 June 1997 (06.06.97)	US	60/057,654	5 September 1997 (05.09.97)	US
60/048,899	6 June 1997 (06.06.97)	US	60/057,646	5 September 1997 (05.09.97)	US
60/048,877	6 June 1997 (06.06.97)	US	60/057,662	5 September 1997 (05.09.97)	US
60/048,963	6 June 1997 (06.06.97)	US	60/057,650	5 September 1997 (05.09.97)	US
			60/057,661	5 September 1997 (05.09.97)	US
			60/057,647	5 September 1997 (05.09.97)	US
			60/070,923	18 December 1997 (18.12.97)	US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

207 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or 10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum 15 (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

20 Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or 25 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes 30 encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying 35 and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, 5 and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

10

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original 15 environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce 20 a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence 30 of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X 35 was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA;
25 followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include
30 Denhardt's reagent, BLOTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such
35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5 The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and 10 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability 15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

20 The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, 25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be 30 branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a 35 nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins

- 5 such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990);
10 Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

- 15 "A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present
20 invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

25 **Polynucleotides and Polypeptides of the Invention**

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in melanocytes and, to a lesser extent, in testes, ovary, kidney and other tissues.

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, disorders of neural crest derived cells including pigmentation defects, melanoma, reproductive organ defects, and defects of the kidney. Similarly,
35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin,

- reproductive, and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to
5 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating disorders that arise from alterations in the number or fate of neural crest derived cells including cancers such as melanoma and
10 defects of the developing reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in infant brain and fetal lung.

- Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders of the brain or lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
20 of the above tissues or cells, particularly of the central nervous and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
25 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating or diagnosing disorders associated with abnormal proliferation of cells in the Central nervous system and developing lung.
30

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in breast lymph node and to a lesser extent in ovarian cancer and chondrosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune responses such as inflammation or immune surveillance for

tumors. This gene may be important for inflammatory responses associated with tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 236 as residues: Lys-45 to Val-50, Lys-69 to Arg-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune responses including those associated with tumor-induced inflammation.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in T-cells and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases involving T-cells such as inflammation, autoimmunity, and cancers including T-cell lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of T-cells and other cells of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating T-cell based disorders such as inflammatory diseases, autoimmune disease and tumors including T-cell lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, autoimmunity, infection, or disorders involving activation of monocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 238 as residues: Asp-19 to Arg-31.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing or treating diseases that result in activation of monocytes including infections, inflammatory responses or autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with terminal deoxynucleotidyltransferase which is thought to be important in catalyzing the elongation of oligo- or polydeoxynucleotide chains.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly those of the blood such as leukemia and deficiencies in neutrophils such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to terminal deoxynucleotidyltransferase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and differential diagnosis of acute leukemia's. Alternatively, this gene may function in the proliferation of neutrophils and be useful as a treatment for neutropenia, for example, following neutropenia as a result of chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

10 The contig exhibits a reasonable homology to the human chorionic gonadotropic (HCG) analogue-GT beta-subunit as disclosed in U.S. Patent No. 5,508,261 and PCT Publication No. WO 92/22568. There is a high degree of conservation of the structurally important cysteine residues in these identities.

15 This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are 20 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or 25 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune 30 system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

35 This gene is expressed primarily in IL-1- and LPS-induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 241 as residues: Ser-14 to Pro-22, Leu-43 to Val-53.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

- This gene is expressed primarily in IL-1 and LPS induced neutrophils. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 242 as residues: Tyr-22 to His-35.
- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth

factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 10 not limited to, immune dysfunctions including cancer of the T lymphocytes and autoimmune disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at 15 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune disorders particularly of T-cell origin and may act as a growth factor for particular subsets of T-cells such as CD4 positive cells which would make this a useful therapeutic for the treatment of HIV and other immune compromising illnesses.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of many developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected 35 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor or differentiation factor for particular cell types in the developing fetus and may be useful in replacement or other types of therapy in cases where the gene is expressed aberrantly.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

- This gene is expressed primarily in T-cells and to a lesser extent in tumor tissue including glioblastoma, meningioma, and Wilm's tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system including autoimmune conditions such as rheumatoid arthritis, inflammatory disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 245 as residues: Thr-9 to Ser-14.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/ modulation of immune function disorders, including rheumatoid arthritis and inflammatory responses.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in placenta and to a lesser extent in fetal liver and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the hematological and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

- 5 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or 10 progenitor cells in the treatment of chemotherapy patients or kidney disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed primarily in stromal cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematopoietic disorders including cancer, neutropenia, anemia, and thrombocytopenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 20 the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 25 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells, in particular following chemotherapy treatment.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

- The translation product of this gene shares sequence homology with epsilon-COP from Bos taurus which is thought to be important as a component of coatomer, a complex of seven proteins, that is the major component of the non-clathrin membrane 35 coat. Preferred polypeptides encoded by this gene comprise the following amino acid sequences:

MAPPAPGPASGGSGEVDELFDVKNAFYIGSYQQCINEAXXVKLSSPERDVERD

VFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMFADYLAHESRRDSIVAELDRE
MSRSXDVTNTFLLMAASIYLHDQNPDAALRALHQGDSLECTAMTVQILLKLD
RLDLARKELKRMQDLDEDATLTQLATAWVSLATGGEKLQDAYYIFQEMADKCS
PTLLLLNGQAACHMAQGRWEAAEGLLQEALDKDSGYPETLVNLIVLSQHLGKP
5 PEVTNRYLSQLKDAHRSHPFKEYQAKENDFDRLVLQYAPSAEAGPELSGP
(SEQ ID NO:458); or RDVERDVFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMF
ADYLAHESRRDSIVAELDREMSRSXDVTNTFLLMAASIYLHDQNPDAALRALH
QGDSLECTAMTVQILLKLDRLARKELKRMQDLDEDATLTQLATAWVSLATG
GEKLQDAYYIFQEMADKCSPTLLLLNGQAACHMAQGRWEAAEGLLQEALDKD
10 SGYPETLVNLIVLSQHLGKPPEVTNRYLSQLKDAHRSHPFKEYQAKENDFDRL
VLQYAPSA (SEQ ID NO:459).

This gene is expressed primarily in activated monocytes and T-cells, and to a lesser extent in multiple other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunomodulation, specifically relating to transport problems in these cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to epsilon-COP indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating /diagnosing problems with the cellular transport of proteins that may result in immunologic dysfunction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with an RNA helicase which is thought to be important in polynucleotide metabolism. The translation product of this contig exhibits good homology to the LbeIF4A antigen of Leishmania braziliensis. The LbeIF4A antigen, or immunogenic portions of it, can be used to induce protective immunity against leishmaniasis, specifically *L. donovani*, *L. chagasi*,

L. infantum, L. major, L. braziliensis, L. panamensis, L. tropica and L. guyanensis. It can also be used diagnostically to detect Leishmania infection or to stimulate a cellular and/or humoral immune response or to stimulate the production of interleukin-12.

This gene is expressed primarily in colon cancer and to a lesser extent in
5 pituitary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers particularly of the colon. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 249 as residues: Glu-93 to Ala-98, Gln-150 to Leu-156, Leu-220 to Leu-231, Leu-268 to Arg-273, Val-324 to Pro-341, Arg-372 to Asn-
20 380, Ser-405 to Gly-410, Phe-426 to Ala-433, Glu-458 to Asp-470, Arg-506 to Ser-547.

The tissue distribution and homology to RNA helicase indicates that polynucleotides and polypeptides corresponding to this gene are useful for development of diagnostic tests for colon cancer.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this contig has sequence homology to a cytoplasmic protein that binds specifically to JNK designated the JNK interacting protein-1 or JIP-1 in mice. JIP-1 caused cytoplasmic retention of JNK and inhibition of JNK-regulated
30 gene expression.

This gene is expressed primarily in brain including pituitary cerebellum frontal cortex, fetal brain and to a lesser extent in the kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of the central nervous system disorders including ischemia, epilepsy, Parkinson's disease, and schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, the translation product of this contig may suppress the effects of the JNK signaling pathway on cellular proliferation, including transformation by the

5 Bcr-Abl oncogene. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 250 as residues: Pro-6 to Ser-26, Ala-30 to Asp-41, Gly-55 to Ser-61, Gly-74 to Thr-80, Tyr-117 to Ala-123, Tyr-167 to Asp-172, Ala-212 to Cys-223, Pro-239 to Tyr-244.

10

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for enhanced survival and/or differentiation of neurons as a treatment for neurodegenerative disease.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The translation product of this gene shares sequence homology with a liver stage antigen from a protozoan parasite.

20

This gene is expressed primarily in fetal tissue and to a lesser extent in activated T-cells and other immune cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and diseases of immune function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

30

35

The tissue distribution and homology to a protozoan antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/immune modulation of parasitic infections.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

Preferred polypeptide encoded by this gene comprise the following polypeptide sequences:

MKAIGIEPSLATYHHIIRLFQPGDPLKRSSFIYDIMNELMGKRFSPKD
PDDDKFFQSAMSICSSLRDLELAYQVHGLLKTGDNWKFIGPDQHRNFYYSKFF

10 DLICLMEQIDVTLKWYEDLIPSAVFPHSQTMIHLLQALDVANRLEVIPKIWER
(SEQ ID NO:460); and/or KDSKEYGHTFRSDLREEILMLMARDKHPPELQVAF
ADCAADIKSAYESQPIRQTAQDWPATSLNCIALFLRAGRTQEAWKMLGLFRKH
NKIPRSELLNELMDSAKVSNSPSQAIEVVELASAFSLPICEGLTQRVMSDFAINQ
EQKEALSNLTALTSDSDTDSSSDSDTSEGK (SEQ ID NO:461). Polynucleotides
15 encoding such polypeptides are also provided.

This gene is expressed primarily in stromal and CD34 depleted bone marrow cells and to a lesser extent in tissues of embryonic origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of hematologic origin including cancers and immune dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
25 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 252 as residues: Ser-28 to Gln-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells which may be useful in the treatment of chemotherapy patients
35 suffering from neutropenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

Preferred polypeptide fragments can be found in an alternative open reading frame. These preferred polypeptides comprise the amino acid sequence:

MSSDNESDIEDEDLKLELRRRLRDKHLKEIQDLQSRQKHEIESLYTKLGKVPPAVI

- 5 IPPAAPLSGRRRPTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQQTL
HPPGNIPESGQNQLLQPLKPSPSSDNLYSAFTSDGAISVPSLSAPGQGTTSNTV
GATVNSQAAQAQPPAMTSSRKGTFTDDLHKLVDNWARDAMNLSGRRGSKGH
MNYEGPGMARKFSAPGQLCISMTSNLGGsapisaasatSLGHFTKSMCPCPQQY
GFPATPFGAQWSGTGGPAPQPLGFQPVGTASLQNFNISNLQKSISNPPGSNL
10 RTT (SEQ ID NO:462); IQDLQSRQKHEIESLYTKLGKVPPAVIIPPAAPLSGRRR
PTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQQTLHPPGNIPESGQN
QLLQPLKPSPSSDNLYSAFTSDGAISVPSLSAPGQGTTS (SEQ ID NO:463);
TSDGAISVPSLSAPGQGTTSNTVGATVNSQAAQAQPPAMTSSRKGTFTDDLH
(SEQ ID NO:464); KGHMNYEGPGMARKFSAPGQLCISMTSNLGGsapisaas
15 ATSLGHFTK (SEQ ID NO:465); QPLKPSPSSDNLYSAFTSDGAISVPSLSAPG
(SEQ ID NO:466). Also preferred are polynucleotide fragments encoding these
polypeptide fragments.

This gene is expressed in fetal liver and tissues associated with the CNS.

- Therefore, polynucleotides and polypeptides of the invention are useful as
20 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, liver and CNS diseases. Similarly, polypeptides and antibodies directed
to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
25 tissues or cells, particularly of the liver and CNS, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
30 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 253 as residues:
Gln-26 to Lys-34.

- The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for diagnosis and treatment for liver diseases such
35 as hepatocellular carcinomas and diseases of the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

In an alternative reading frame, this gene shows sequence homology to two recently cloned genes, karyopherin beta 3 and Ran_GTP binding protein 5. (See Accession Nos. gi|2102696 and gnl|PID|e328731.) The Ran_GTP binding protein is related to importin-beta, the key mediator of nuclear localization signal (NLS)-dependent nuclear transport. Based on homology, it is likely that this gene may activity similar to the RAN_GTP binding protein. Preferred polypeptide fragments comprise the amino acid sequence: VRVAAAESMXLLCAXVRGPEYLTQMWHFMCDALIKA IGTEPDSDLVSEIMHSFAK (SEQ ID NO:467). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in thymus tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in prostate and osteoclastoma tissues. Preferred polypeptide fragments also comprise the amino acid sequence: MEINNQNCFIVIDLVRTVMENGVEGLLIFGAFLPESWLVRCSSSEPPKALLLIL AHSQKRRLDGWSFIRHLRVHYCVSLTIHFS (SEQ ID NO:468). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone and prostate diseases, and cancers, particularly of the bone and prostate. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and prostate systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 255 as residues: Met-1 to Ser-11.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for bone and prostate disorders, especially cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

15 This gene shares sequence homology with the FK506-binding protein (FKBP-13) family, a known cytosolic receptor for the immunosuppressants. Recently, another group has cloned a very similar gene, recognizing the homology to FK506-binding protein family, calling their gene FKBP23. (See Accession No. 2827255.)

20 This gene is expressed primarily in lymphoid tissues.
Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample, especially for those susceptible to immune suppressant therapies and for diagnosis of diseases and conditions, which include, but are not limited to, immune suppressant disorders. Similarly, polypeptides and antibodies directed to these
25 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 256 as residues: Ala-19 to Val-31, Arg-38 to Gly-49, Ala-61 to Lys-66, Tyr-68 to Pro-78, Gly-116 to Ala-121, Asp-154 to
35 Ser-162, Glu-173 to Gln-186, Phe-194 to Gly-203, Pro-207 to Val-212.

The tissue distribution and homology to FKBP-12 and -13 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune suppressant disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in the brain and in the retina. This gene maps to chromosome 8, and therefore can be used in linkage analysis as a marker for chromosome 8.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and ocular associated disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of
- 15 disorders of the above tissues or cells, particularly of the disorders of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
- 20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 257 as residues: Cys-34 to Asp-40.

- The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma. Expression in the brain indicates a role in the is useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

- This gene shows sequence homology to a newly identified class of proteins expressed in the nervous system, called stathmin family. (See Accession No. 2585991; 35 see also Eur. J. Biochem. 248 (3), 794-806 (1997).) The stathmin family appears to be an ubiquitous phosphoprotein involved as a relay integrating various intracellular signaling pathways. These pathways affect cell proliferation and differentiation.

Preferred polypeptide fragments comprise the amino acid sequence:

QDKHAEVRKNKELKEEASR (SEQ ID NO:469); QQDLSPWAAPVGCPPLXXASX
TCHXLPLSGCLRRQSXSLPVVAXLCFWFSCPLASLFVPGQPCVTCPFPSSLFQD
KHAEEVRKNKELKEEASR (SEQ ID NO:470). Also preferred are the

5 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

10 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntintons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 26

The polynucleotide sequence of this gene contains a domain similar to a Flt3 ligand peptide. Preferred polypeptide fragments comprise the amino acid sequence:

PTRCCTTQPCRSSARRPCWVPMVPSPEGREXQPTCPS (SEQ ID NO:471). Thus, this gene may have activity as binding to Flt3 receptors, a process known to promote
30 angiogenesis and/or lymphangiogenesis.

This gene is expressed in human tonsil, and to a lesser extent in teratocarcinoma, placenta, colon carcinoma, and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the tonsil, as well as cancers, such as colon, reproductive, and kidney cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful

in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tonsils, colon, reproductive organs, and kidneys, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 259 as residues:

10 Pro-22 to Glu-33.

The tissue distribution in tonsil and several cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the tonsil or colon, such as tonsillitis, inflammatory diseases involving nose and paranasal sinuses, especially during the 15 infection of influenza, adenoviruses, parainfluenza, rhinoviruses. The gene may also be useful in the diagnosis and treatment of neoplasms of nasopharynx or colon origins.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In an alternative reading frame exists a large open reading frame that encodes a 20 preferred polypeptide. Preferred polypeptide fragments comprise the amino acid sequence:

MKRSLNENSARSTAGCLPVPLFNQKKRNRQPLTSNPLKDDSGISTPSDNYDFP
PLPTDW AWEAVNPEXAPVMKTVDTGQIPHVSRSPLRSQDSVFNSIQSNTGRSQ
GGWSYRDGNKNTSLKTWXKNDFKPQCKRTNLVANDGKNSCPMSSGAQQQK
25 QLRTPEPPNLSRNKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNFQQNQY
KXQMLDDIPEDNTLKETSLYQLQFKEKASSLRIISA VIESMKYWREHAQKTVLL
FEVLA VLDSA VTPGPYYSKTFLMRDGKNTLPCVFYEIDREL PRLIRGRVHRCVG
NYDQKKNIFQCVSVRPASVSEQKTFQAFVKIADVEMQYYINVNET (SEQ ID
30 NO:472); SQDSVFNSIQSNTGRSQGGWSYRDGNKNTSLKTWXKNDFKPQCKR
(SEQ ID NO:473); NKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNF (SEQ ID
NO:474); SSLRIISA VIESMKYWREHAQKTVLLFEVLA VLDSA VTPGPYYSKTFLM
(SEQ ID NO:475); and PRLIRGRVHRCVGNYDQKKNIFQCVSVRPASVSEQKT
FQAFV (SEQ ID NO:476).

This gene is expressed primarily in human testes.
35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, male reproductive disorders, including cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system,

5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

10 disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a hormone with reproductive or other systemic functions; contraceptive development; male infertility of testicular causes, such as Kleinfelter's syndrome, varicocele, orchitis; male sexual dysfunctions; testicular neoplasms; and inflammatory disorders such as epididymitis.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases relating to T cells, as well as cancer in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

25

30

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders. Moreover, since the gene was isolated from an apoptotic cell and based on the understanding of the relationship of apoptosis and cancer, it is likely that this gene may play a role in the genesis of cancer.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
- 10 tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level
- 15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal diseases.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with C44C1.2 gene product of *Caenorhabditis elegans* with unknown function. Preferred polypeptide fragments comprise the amino acid sequence:

- GVRPCVCGRPASLTCSPLDPEVGPYCDTPTMRTLNFNLLWLALACSPVHTTLSK
- 25 SDAKKAASKTLLEKSQFSDKPVQDRGLVVTDLKAESVVLEHRSYCSAKARDRH
FAGDVLYVTPWNSHGYDVTKVFGSKFTQISPVWLQLKRRGREMFETVGLHD
VDQGWMRAVRKHAKGLHIVPRLLFEDWTYDDFRNVLDSEDEIEELSKTVVQVA
KNQHFDGFVVEVWNQLLSQKRVGLIHMMLTHLAEALHQARLLALLVPPAIPGT
DQLGMFTHKEFEQLAPVLDGFSLMTYDYSTAHQPGPNAPLSWVRACVQVLDP
- 30 KXXWRTKSSWGSTMWXRXXPXDARXPVVGXRXIQLKDHXPRMVLDSK
PQ (SEQ ID NO:477); TCSPLDPEVGPYCDTPTMRTLNFNLLWLALACSPVHTTLS
(SEQ ID NO:478); LVVTDLKAESVVLEHRSYCSAKARDRHFAGDVLYVTPW
NSHGYDVTKVFGSKF (SEQ ID NO:479); REMFEVTGLHDVDQGWMRAVRK
HAKGLHIVPRLLFEDWTYDDFRNVLDSEDE (SEQ ID NO:480); HFDGFVVEVW
- 35 NQLLSQKRVGLIHMMLTHLAEALHQARLLALLVPPAIPGT DQLGM (SEQ ID
NO:481); DGFSLMTYDYSTAHQPGPNAPLSWVRACVQVLDPKXXWRTKSSW
GST (SEQ ID NO:482). Also preferred are polynucleotide fragments encoding these

polypeptide fragments. This gene maps to human chromosome 11, and therefore is useful in linkage analysis as a marker for chromosome 11.

This gene is expressed primarily in human T cells and to a lesser extent in human colon carcinoma.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for
- 10 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
- 15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 263 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.
- 20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders and gastrointestinal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

- 25 The translation product of this gene shares sequence homology with Ribosomal protein L11 of *Caenorhabditis elegans*. (See Accession No. 156201.) Preferred polypeptide fragments comprise the amino acid sequence:
- ERGV SINQFC KEFNERTK DIKEGI PLPTKILVK PDRTFEIKIG QPTV SYFL KAAAG
IEKG ARQTG KEVAGL VTLKH VYEIARIKA QDEA FALQ DVPLSS VVR SIIGSARSL
- 30 GIRVV KDL SSEE LAAF QKERA IFLAAQ KEAD LAAQ EEA AKK (SEQ ID NO:483).
Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in human embryo tissue and to a lesser extent in human epithelioid sarcoma and other tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 35 reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development disorders and epithelial cell cancer. Similarly, polypeptides and antibodies

directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic and epithelial cell systems, expression of this gene at significantly higher or lower levels may be routinely detected

- 5 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID

10 NO: 264 as residues: Lys-34 to Gly-40.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental disorders and epithelial cancer.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

This gene is expressed primarily in resting T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

- 20 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders of immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is believed to reside on chromosome 1. Accordingly, polynucleotides derived from this gene are useful in linkage analysis as chromosome 1 markers.

This gene is expressed primarily in prostate and to a lesser extent in soares adult brain, human umbilical vein endothelial cells, and amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary system and nervous system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of disorders of the urinary and nervous systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene shares sequence homology with R05G6.4 gene product. (See Accession No. gil1326338.) This gene also shares sequence homology with the cyclophilin-like protein CyP-60. (See Accession No. 1199598, see also Biochem. J. 314 (1), 313-319 (1996).) Preferred polypeptide fragments comprise the amino acid sequence:
AVYTYHEKKKDTAASGYGTQNIRLSRDAVKDFDCCCLSLQPCHDPVVTPDGYL
YEREALYEYLHQKKEIARQMKAYEKQRGTRREEQKELQRAASQDHVRGFLEKE
SAIVSRP LNPFTAKALSGTSPDDVQPGPSVGPPSKDKDKVLPSFWIPSLTPEAK
ATKLEKPSRTVTCPMSGKPLRMSDLTPVHFTPPLDSSVDRVGLITRSERYVCATV
RDSLNSATPCAVLRPSGAVVTLECVEKLIRKDMVDPVTGDKLTDRIIVLQRGT
(SEQ ID NO:484); YLYEREALYEYLHQKKEIARQMKAYEKQRGTRREEQKELQ
RAASQDHVRGFLE (SEQ ID NO:485); and FTAKALSGTSPDDVQPGPSVGPP
SKDKDKVLPSFWIPSLTPEAKATKLEKPSRTVTCPMSGKPL (SEQ ID NO:486).

Also preferred are polynucleotide fragments that encode these polypeptide fragments.

This gene is expressed primarily in human testis and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders and in particular testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system. Expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the male reproductive system and in particular of testicular cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

The translation product of this gene shares sequence homology with Lpe5p of *Saccharomyces cerevisiae* which is thought to be important in the metabolism of phospholipids.

This gene is expressed primarily in liver and brain. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 268 as residues: Pro-14 to Leu-20, Lys-28 to Asn-38, Arg-109 to Arg-114, Lys-119 to Asn-124, Glu-152 to Leu-157, Pro-172 to Val-180.

The tissue distribution and homology to Lpe5p of *Saccharomyces cerevisiae* indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of metabolic and nervous disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene shares sequence homology with the nuclear ribonucleoprotein U (HNRNP U), encoded by *C. elegans* (See Accession gil1703576.) Preferred polypeptide fragments comprise the amino acid sequence:

- 5 MDTSENRPENDVPEPPMPIADQVSNDRPEGSVEDEEKKESSLPKSFKRKISVV
SATKGVPAGNSDTEGGQPGRKRRWGASTATTQKKPSISITTESLKSILPDIKPL
AGQEAVVDLHADDSRISEDETERNGDDGTHDKGLKICRTVTQVVPAEGQENGQ
REEEEEEEKEPEAEPVPVPPQSVVALPPAEHEVKVTLGDTLRRSISQQKSGV
SITIDDPVRTAQVPSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAFWI
- 10 DKIKSHCFVTYSTVEEAVATRTALHGKVWPQSNSPKFLCADYAEQDELDYHRGL
LVDRPSETKTEEQGIPRPLHPPPPPQVQPPQHPRAEQREQERAVERQWAERERE
MERRERTRSEREWDRDKVREGPRSRSRSRXRRRKERAKSKEKKSEKKEKAQE
EPPAKLDDLFRKTKAAPCITYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEE
EQKEREKEAERERNRQLEREKRREHSRERDRERERERDRGDRDRDRERDRE
- 15 RGRERDRRDTKRHSRSRSRSTPVRDRGGR (SEQ ID NO:488). Also preferred are the polynucleotide fragments encoding this polypeptide fragments.

This gene is expressed primarily in epididymus.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the male reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of male reproductive disorders.

35 **FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory diseases and reproductive disorders. Similarly,

- 5 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from 10 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene shares sequence homology with human opsonin protein P35 fragment. (See Accession No. R94181.) The opsonin protein activates the phagocytosis of pathogenic microbes by phagocytic cells. Preferred polypeptide fragments comprise the amino acid sequence: GCDSCPPLPRAFAQDTQAEGECSSRAERADMCPDAP PSQEVPPEGPGAAP (SEQ ID NO:489). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed in immune-related tissues such as thymus, macrophage, T cells and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and infectious disease. Similarly, polypeptides and 30 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and infectious disease, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from 35 an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 271 as residues: Lys-9 to Arg-14, Met-38 to Asp-51.

The tissue distribution indicates that polynucleotides and polypeptides
5 corresponding to this gene are useful for diagnosis and treatment of immune disorders,
as well as the treatment and/or diagnosis of infectious disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

The translation product of this gene shares sequence homology with alpha-2
10 type I collagen which is thought to be important in tissue repair. (See, e.g., 211607.)
Preferred polypeptide fragments comprise the amino acid sequence: PQLPSCGRPW
PGTASVFQSHTQGPREDPDPCRAQGSAGTHCPISLSPPRQ (SEQ ID NO:490).
Also preferred are the polynucleotide sequences encoding these polypeptide sequences.

This gene is expressed primarily in the brain and to a lesser extent in the kidney
15 and thymus

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, brain, kidney, and immune disorders. Similarly, polypeptides and
20 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the brain, kidney, and immune disorders,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution and homology to alpha-2 type I collagen indicates that
30 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
and treatment of tissue repair, and brain, kidney, immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

The translation product of this gene shares sequence homology with mini-
35 collagen which is thought to be important in tissue repair tumor metastasis. (See
Accession No. gnl|PID|d1006976.) Preferred polypeptide fragments comprise the
amino acid sequence: PGFRGPGSGSLGCSFFPRSLGRVLPPGCQRPGAHAD

SSPPPTP (SEQ ID NO:491). Also preferred are polynucleotides encoding this polypeptide fragment.

This gene is expressed in ovarian cancer and to a lesser extent in dendritic cells and smooth muscle.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumor metastasis and tissue repair. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor metastasis and tissue repair, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 273 as residues: Asn-2 to His-11.

The tissue distribution and homology to mini-collegen gene indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumor metastasis and tissue repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene shares sequence homology with the HIV TAT protein. (See
25 Accession No. 328416.) Preferred polypeptide fragments comprise the amino acid sequence: EDLKKPDPASLRAASCSEGKKRKACKNCTCGLAEELEKEK SREQMSSQPKSACGN CYLGDAFRCASC P YLGMPAFKPGEKVLLS (SEQ ID NO:492); EDLKKPDPASLRAASCSEGKKRKACKNCTCGLAEELEKEK SREQMSSQPKSACGN CYLGDAFRCASC P YLGMPAFKPGEKVLLSDSNLHD
30 (SEQ ID NO:493); CGNCYLGDAFRCASC P YLGMPAFKPGEKVLLSDS (SEQ ID NO:494); SCGEGKKRKACKNCTCGLAEELEKE (SEQ ID NO:495); SQPKSAC GNCYLGDAFRCASC (SEQ ID NO:496); and REAGQNSERQYVS LSRD (SEQ ID NO:497). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

35 This gene is expressed primarily in the infant brain and to a lesser extent in the breast and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, testes and breast disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, testes and breast disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 274 as residues: Pro-7 to Val-15.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of brain, testes and breast, and other related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in the infant brain, human cerebellum, and to a lesser extent in medulloblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain related disorders and medulloblastoma and other brain cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain related disorders and brain cancers, including medulloblastoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 275 as residues: Thr-41 to Glu-47.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain related disorders, brain cancers, and medulloblastoma.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The translation product of this gene shares sequence homology with a phosphotyrosine-independent ligand for the lck SH2 domain which is thought to be important in signal transduction related to phosphotyrosine-independent ligand for the lck SH2 domain. (See Accession No. gil1184951.) Preferred polypeptide fragments

10 comprise the amino acid sequence: ESSGQARTLADPGPGWPRQQGMCFGSLT GLSTTPHGFLTVSAEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAAL DTIQYSKH (SEQ ID NO:498). Also preferred are polynucleotide fragments encoding this polypeptide fragment. It is likely that this gene is a new member of a family of phosphotyrosine-independent ligands for the lck SH2 domains.

15 This gene is expressed primarily in the placenta and to a lesser extent in endothelial cells and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 20 not limited to, reproductive, cardiovascular, immune, and infectious diseases.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, reproductive, and immune system, and infectious diseases, expression 25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the 30 disorder.

The tissue distribution and homology to a phosphotyrosine-independent ligand for the lck SH2 domain indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cardiovascular, reproductive, and immune system diseases, as well as infectious diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

This gene is expressed primarily in the fetal brain, cerebellum and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal cell related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell related disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 277 as residues: Thr-20 to Gly-28.

The tissue distribution and homology to proline-rich protein genes indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with precerebellin of human, which is thought to be important in synaptic physiology. (See Accession No. gil180251.) It has been observed that cerebellin-like immunoreactivity is associated with Purkinje cell postsynaptic structures. Thus, it is likely that this gene also have synaptic activity. Preferred polypeptide fragments comprise the amino acid sequence: QEGSEPVLLGECLVVCEPGRAAAGGPGGAALGEAPPGRVAFXAV RSHHHEPAGETGNGTSGAIYFDQVLVNEGGGFDRASGSFVAPVRGVYSFRFH VVKVYNRQTVQVSLMLNTWPVISAFANDPDVTREAATSSVLLPLDPGDRVSLR LRRGXSTGW (SEQ ID NO:499). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in cerebellum and infant brain. By Northern analysis, a single transcript of 2.4 kb was observed in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, neuronal cell signal transduction and synaptic physiology. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell

- 5 signal transduction and synaptic physiology expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue
10 or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene or gene family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed in fetal liver and spleen, and to a lesser extent in bone marrow, umbilical vein, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

- 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the immune system, particularly hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis
25 and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 279 as residues: Asp-30 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopietic and immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares sequence homology with a 12 kD nucleic acid binding protein of Feline calcivirus which is thought to be important in viral replication. (See Accession No. 59264)

- 5 This gene is expressed primarily in human cardiomyopathy and to a lesser extent in T helper cells, fetal brain and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 10 not limited to, cardiomyopathy as well as viral infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain 15 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID 20 NO: 280 as residues: Trp-20 to Cys-26.

The tissue distribution in cardiomyopathy and homology to viral 12 kD nucleic acid binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cardiomyopathy, including those caused by ischemic, hypertensive, congenital, valvular, or pericardial abnormalities.

- 25 The gene expression pattern may be the consequence or the cause for these conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with tumor necrosis factor related gene product which is thought to be important in tumor necrosis, 30 bacterial and viral infection, immune diseases and immunoreactions.

This gene is expressed primarily in colon and to a lesser extent in ovarian and breast cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary or breast origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 5 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Tumor necrosis factors indicates that 10 polynucleotides and polypeptides corresponding to this gene are useful for intervention of cancers of colon, ovary and breast origins, because TNF family members are known to be involved in the tumor development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

15 The translation product of this gene shares sequence homology with mucins, such as epithelial mucin, which is thought to be important in extracellular matrix functions such as protection, lubrication and cell adhesion (See for example Accession No. R68002). Preferred polypeptide fragments comprise the following amino acid sequence: PRSRPALARPGQRPPSHSATSGVLRPRKKPDP (SEQ ID NO:500).
20 Also preferred are polynucleotide fragments encoding these polypeptide fragments. Moreover, this gene maps to chromosome 22q11.2-qter, and therefore, can be used as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in corpus colosum.

Therefore, polynucleotides and polypeptides of the invention are useful as 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors, especially of corpus colosum, as well as metastatic lesions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell 30 type(s). For a number of disorders of the above tissues or cells, particularly of the corpus colosum and other solid tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, 35 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to mucins indicates that polynucleotides and polypeptides corresponding to this gene are useful for serum tumor markers or immunotherapy targets because tumor cells have greatly elevated level of mucin expression and shed the molecules into the epithelial tissues.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in CD34 depleted buffy coat cord blood and primary dendritic cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disorders and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34 depleted buffy coat cord blood and primary dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders. Secreted or cell surface proteins in the above tissue distribution often are involved in cell activation (e.g. cytokines) or molecules involved in cell surface activation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

- The translation product of this gene shares sequence homology with Interferon induced 1-8 gene encoded polypeptide which is thought to be important in binding to retroviral rev responsive element. Preferred polypeptide fragment comprise the following amino acid sequences: MTLITPSXKLTFXKGNKSWSSRACSSTLVDP (SEQ ID NO:501). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in CD34 positive cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, retroviral infection, such as AIDS, and other immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 284 as residues: Gln-51 to Trp-62.

The tissue distribution and homology to interferon induced gene 1-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of retroviral infection including HIV. The factor may be involved in viral stability or viral entry into the cells. Alternatively, the virus/factor complex may elicit the cellular immune reaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene shares sequence homology to immunoglobulin lambda chain (See Accession No. 2865484). Therefore it is likely that this gene has activity similar to an immunoglobulin lambda chain. Preferred polypeptide fragments comprise the following amino acid sequence: GHPSPALSIAPS DGSQQLPCDEV PYGEAHVTRY CKKPLTNS HLETEAQSSSL (SEQ ID NO:502). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma and other immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 285 as residues: Pro-27 to Thr-32.

- The tissue distribution in Hodgkin's lymphoma and the sequence homology
- 5 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune
- 10 functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

- This gene has extensive homology to cDNA for Homo sapiens mRNA for the
- 15 ISLR gene(See Accession No. AB003184). This protein is considered to be a new member of the Ig superfamily and contains a leucine-rich repeat (LRR) with conserved flanking sequences and a C2-type immunoglobulin (Ig)-like domain. These domains are important for protein-protein interaction or cell adhesion, and therefore it is possible that the novel protein ISLR may also interact with other proteins or cells. The ISLR gene
- 20 was mapped on human chromosome 15q23-q24 by fluorescence in situ hybridization (See Medline Article No. 97468140). Homology to the ISLR gene has been confirmed by another independent group as well (See Accession No. Hs.102171)

- This gene is expressed in a number of tissues including human retina, heart, skeletal muscle, prostate, ovary, small intestine, thyroid, adrenal cortex, testis,
- 25 stomach, spinal cord, fetal lung and fetal kidney tissues, colon, tonsil and stomach cancer, and to a lesser extent in endometrial stromal cells treated with estradiol, breast tissue, synovium, lymphoma, and number of other tumors.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
- 30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary and breast origins. However, due to the wide range of expression in various tissues, protein may play a vital role in the development of cancer in other tissues as well, not just those mentioned above. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an

- 5 individual not having the disorder. Additionally, this gene maps to chromosome 15q23-q24, and therefore, can be used as a marker in linkage analysis for chromosome 15.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where
10 expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

- 15 Gene has homology to multidrug resistance gene 1 (See Accession No. P06795). Preferred polynucleotide fragments comprise the following sequence:
GCTTCGTGTCCAACCCCTTGCCTCGCCTGTGTGCCAGGCCAGTCCC
CCACGCTCGCGTTCCCTGTAGTGCTCACAGGTCCCAGCACCGATGGCA
TTCCCTTGCCCTGAGTCTGCAGCGGGTCCCTTTGTGCTTCCTCCCCCTCA
20 GGTAGCCTCTCTCCCCCTGGGCCACTCCGGGGTGAGGGGGTACCCCTT
CCCAGTGTAAAAATTCCCTGTGGGGCTCACCCAAAGTATTAAAAGTAGCTTT
GTAA (SEQ ID NO:503). Also preferred are polypeptide fragments encoded by these polynucleotide fragments.

This gene is expressed primarily in lung, esophagus, leukemia (Jurkat cells) and
25 breast cancers and to a lesser extent in macrophages treated with GM-CSF fetal tissues and wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are
30 not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower levels may be routinely detected
35 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, due to the high expression level in lung tissue and the proposed function of the multidrug resistance protein 1 gene as the efflux pump responsible for low-drug accumulation in multidrug-resistant cells, protein as well mutants thereof,
5 may also be beneficial as a target for gene therapy, particularly for the chronic patient. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 287 as residues: Met-1 to Lys-16.

The tissue distribution in wide range of cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of
10 cells in active proliferation, such as cancers. The gene products may be used for cancer markers or immunotherapy target.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene maps to the X chromosome.

15 This gene is expressed primarily in the brain and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, neurodegenerative disease states and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders, including sex-linked disorders, of the above tissues or cells, particularly of the neurological, developmental systems, and cardiovascular system,
25 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder. Moreover, this gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for this chromosome.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
35 Disease, Huntington's Disease, Klinefelter's, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

- 5 The translation product of this gene shares sequence homology with paxillin which is thought to be important in mediating signal transduction from growth factor receptors to the cytoskeleton. Preferred polynucleotide fragments comprise the following sequence: TGGCTCACTGTCTTACAATCACTGCTGTGGAATCATGA
TACCACTTTAGCTTTGCATCTCCTTCAGTGTATTTGTITTCAGAGG
10 AAGTAGATTAACTGGACAACTTGAGTACTGACATCATTGATAAATAAACT
GGCTGTGGTTCAA (SEQ ID NO:506). Also preferred are polypeptide fragments encoded by these polynucleotide fragments. More preferably, polypeptide fragments comprise the amino acid sequence: LDELMAHLTEMQAKVAVRAD
AGKKHLPDKQDHKASLDSMLGGLEQELQDLGIATVPKGHCASCQKPIAGKVI
15 HALGQSWHPEHFVCTHCKEEIGSSPFFERSGLXYCPNDYHQLFSPRCAYCAAP
ILDKVLTAMNQTWHPEHFFCSHCGEVFGAEGFHEKDKKPYCRKDFLAMFSPK
CGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCELHYH
HRRGTLCHGCGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFREQNDKY
CQPCFNKLF (SEQ ID NO:507); KASLDSMLGGLEQELQDLGIATVPKGHC
20 ASCQKPIAGKVIHAL (SEQ ID NO:508); CPNDYHQLFSPRCAYCAAPILDKVL
TAMNQTWHPEHFFCSHCGEVFGAEG (SEQ ID NO:509); DKKPYCRKDFLAM
FSPKCGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCE
L (SEQ ID NO:510); CGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFRE
QNDKTYCQ (SEQ ID NO:511). Polynucleotide fragments encoding these preferred
25 polypeptide fragments are also contemplated.

This gene is expressed primarily in brain, and to a lesser extent in the developing embryo.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disease states and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and
35 nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, since this gene shares homology with a gene that maps to chromosome 11, (See Accession No.T87404), gene as well as its 5 translated product may be used for linkage analysis on chromosome 11.

The tissue distribution and homology to paxillin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and or detection of disease states associated with abnormal signal transduction in brain and/or the developing embryo. This would include treatment or detection of neurodegenerative 10 disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in fetal spleen, brain, and to a lesser extent in six week old embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, neurological disorders, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of 25 the immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue 30 or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 290 as residues: Arg-28 to Gly-34.

The expression of this gene in fetal spleen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune 35 disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. In addition the expression of this gene in the early embryo, indicates a key role in embryo development and hence the gene or gene product could be used in the treatment and or detection of embryonic development defects. This would include

treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntintons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with the gene disrupted in the neurodegenerative disease dentatorubal-pallidoluysian atrophy. Moreover a long open reading fame exists in an alternative frame. Preferred polypeptide fragments

10 comprise the following:

MGSSQSVEIPGGGTEGYHVLRVQENSPGHRAGLEPFFDFIVSINGSRLNKDND
TLKDLLKXNVEKPVKMLIYSSKTLELRETSVTPSNLWGGQGLLGVSIRFCSD
GANENVWHVLEVESNSPAALAGLRPHSDYIIGADTVMNESEDLFSLIETHEAKP
LKLYVYNTTDNCREVIITPNSAWGGEGLCGIGYGYLHRIPTRPFEEGKKIS
15 LPGQMAGTPITPLKDGFTEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVSS
VLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLNLPA
PHIMPGVGLPELVNPGLPPLPSMPPRNLPGIAPLPLPSEFLPSFPLVPESSAASS
GELLSSLPPTSNAPSDPATTTAKADAASSLTVDVTPTAKAPTTVEDRVDSTPV
SEKPVSAAVDANASESP (SEQ ID NO:512); SVEIPGGGTEGYHVLRVQENSPGH
20 RAGLEPFFDFIVSINGSRLNKDNDTLKDLLKXNVEKPVKMLIYSSKTLELRETS
VTPSNLWGGQGLLGVSIRFCSDGANENVWH (SEQ ID NO:513); ESNSPAA
LAGLRPHSDYIIGADTVMNESEDLFSLIETHEAKPLKLKYVYNTTDNCREVIITP
NSAWGGEGLCGIGYGYLHRIPTRPFEEGKKISLPGQMAGTPITPLKDGFTEV
QLSSVNPPSLSPPGTTGIEQSLTG LSISS (SEQ ID NO:514); RIPTRPFEEGKKI
25 SLPGQMAGTPITPLKDGFTEVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS
SVLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLNLNP
APHIMPGVGLPELVNPGLPPLPSMPPRN (SEQ ID NO:516); PGLPPLPSMPPRN
LPGIAPLPLPSEFLPSFPLVPESSAASSGELLSSLPPTSNAPSDPATTTAKADA
SSLTVDTPTAKAPTTVEDRVDSTPVSEKPVSAAVDAN (SEQ ID NO:517).

30 This gene is expressed primarily in prostate cancer, and to a lesser extent in the pineal glands and in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 35 not limited to, neurological conditions and pulmonary disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the nervous, pulmonary, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or 5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 291 as residues: Asn-9 to Leu-14.

The abundance of this gene in the pineal gland and its homology to a gene 10 disrupted in the neurodegenerative disease state Dentatorubral-pallidoluysian atrophy indicates that this gene may be useful in the treatment and/or detection of other neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. The abundance of this gene in fetal 15 lung would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung; that it may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis; and thus the gene or the gene protein encoded by the gene could be used in the detection and/or treatment of these 20 pulmonary disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 30 the above tissues or cells, particularly of the developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the 35 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene primarily in the embryo, indicates the gene plays a key role in embryo development and that the gene or the protein encoded by the gene could be used in the treatment and or detection of developmental defects in the embryo or in infants.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene displays homology to nestin, an intermediate filament protein, the expression of which correlates with the proliferation of Central Nervous System progenitor cells and that is useful in the identification of brain tumors. This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. AA527348).

- This gene is expressed primarily in kidney and to a lesser extent in brain. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the excretory and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 293 as residues: Thr-128 to Asn-135.

- The tissue distribution and homology to nestin indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and/or treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, its abundance in kidney indicates that it is useful in the treatment and detection of acute renal failure and other disease states associated with the kidney.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 61

Gene shares homology with the latrophilin-related protein 1 precursor as well as the calcium-independent alpha-latrotoxin receptor. Preferred polypeptide fragments

comprise the following amino acid sequence:

IYKVFRHTAGLKPEVSCFENIRSCARXXXXXXXXXWIFGVLVHVVHASVV

TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCC (SEQ ID NO:518);

WIFGVLVHVVHASVVTAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVP

5 C (SEQ ID NO:519). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 2213659) The translation product of this gene shares sequence homology with CD 97, a seven transmembrane bound receptor.

This gene is expressed primarily in infant brain and in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For 15 a number of disorders of the above tissues or cells, particularly of the neurological and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the 20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 294 as residues: Lys-13 to Leu-21.

The tissue distribution of this gene suggest that it may be useful in the detection and/or treatment of neurodegenerative disease states and behavioral disorders such as

25 Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder, while its expression in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma and immunodeficiency diseases.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in fetal liver and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,

5 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 295 as residues: Ser-91 to Lys-98.

10 The tissue distribution of this gene fetal liver and spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma and immunodeficiency diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

15 Gene shares homology with human serum amyloid protein. Preferred polypeptide fragments comprise the following amino acid sequence:
ALTRIPPGDWVINVTAVSFAGKTTARFFHSSPPSLGDQARTDPGHQRRD (SEQ ID NO:520) (See Accession No. W13671). Also preferred are polynucleotide fragments encoding these polypeptide fragments This gene maps to chromosome 9, and
20 therefore, may be used as a marker in linkage analysis for chromosome 9 (See Accession No. AA004342).

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
35

The tissue distribution of this gene in fetal liver-spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma, and immunodeficiency diseases.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 64

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. AA219669).

This gene is expressed specifically in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
- 15 the above tissues or cells, particularly of the neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the
- 20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntintons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

- Gene shares homology with a yeast protein. Preferred polypeptide fragments comprise the following amino acid sequence: LQEVNITLPENSVWYERYKFDIP VFHL (SEQ ID NO:521). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 1332638)

This gene is expressed primarily in fetal tissue (fetus and fetal liver).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver disorders and cancers (e.g. hepatoblastoma). Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected

- 5 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID

10 NO: 298 as residues: Asn-59 to Glu-64.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In 15 addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

- 20 Gene has homology with a B-cell surface antigen which may indicate gene plays a role in the immune response, including, but not limited to disorders and infections of the immune system. Preferred polynucleotide fragments comprise the following sequence: TAGCATGTAGCCAGTCGAATAACNTATAAGGACAAAGTGGAGTC CACGCGTGCAGGCCGTCTAGACTAGTGGATCCCCGGCTGCAGGATTGGC 25 ACGAG (SEQ ID NO:523). Also preferred are polypeptide fragments encoded by these polynucleotide fragments (See Accession No.T94535). Additionally, this gene shares homology with an interferon-gamma receptor. Preferred polypeptide fragments also comprise the following amino acid sequence: MQGSGSQFRACLLCLCFSCPC SPGGPRWNSRQGGRRFPKTCRAISQNLVFKYKTFCPVRYMQPHRSSLCLHFTS 30 YVFILSTW GSLRTYSTD LKKKKNSRGGPVPIRPKS (SEQ ID NO:522); MQGSGSQFRACLLCLCFSCPCSPGGPRWNSRQGGRRFPKTCRAISQNLVFK (SEQ ID NO:524); PVR YMQPHRSSLCLHFTSYVFILSTW GSLRTYSTD LKKKK NSRGGPVPIRPKS (SEQ ID NO:525); and GEEQRDCSLGWRGVGMRATHCQAA RMFVLFLPKYAGL (SEQ ID NO:526). Also preferred are polynucleotide fragments 35 encoding these polypeptide fragments

This gene is expressed primarily in T-cells and gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders and conditions (immunodeficiencies, cancer, leukemia, hematopoiesis). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 299 as residues: 10 Thr-41 to Gly-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune disorders, immuno-suppressive (transplantation) and immunodeficiencies (e.g. AIDS), inflammation and 15 hematopoietic disorders. The expression of this gene in gall bladder would suggest a possible role for this gene product in digestive disorders, particularly of the pancreas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene maps to chromosome 11, and therefore, may be used as a marker in 25 linkage analysis for chromosome 11 (See Accession No. AA011622).

This gene is expressed primarily in a variety of fetal and developmental tissues (e.g. fetal spleen, infant brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, immune or neurological abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing 35 immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 300 as residues: Ser-38 to Ser-43.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for developmental abnormalities or fetal deficiencies. The detection in infant brain would suggest a role in neurological disorders (both developmental and neurodegenerative conditions of the brain and nervous system, behavioral disorders, depression, schizophrenia, Alzheimer's disease, Parkinson's
10 disease, Huntington's disease, mania, dementia). In addition, the detection in spleen would similarly suggest a role in detection and treatment of immunologically mediated disorders (e.g. immunodeficiency, inflammation, cancer, wound healing, tissue repair, hematopoeisis).

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

This gene is expressed primarily in spleen, T-cells, and fetal heart.
Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, immunological deficiencies, including AIDS and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and cardiovascular systems, expression of this gene at significantly higher
25 or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, autoimmune disorders, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. The expression in fetal heart indicates that polynucleotides and
35 polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Gene shares homology with a human collagen protein. Preferred polypeptide fragments comprise the following amino acid sequence:

- 5 MPRKTSKCRQLLCGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPGCXSV
SSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHSKSQGE
GQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGGVKVAATTEREPEFKIK
TGKA (SEQ ID NO:527); CSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPG
CXSVPSSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHS
10 (SEQ ID NO:528); QGEGQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGG
VKVAATTEREPEFKIKTGKA (SEQ ID NO:529) (See Accession No. 124886). Also preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 20 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 302 as residues: 25 Pro-32 to Ser-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

The translation product of this gene shares sequence homology with a chicken single-strand DNA-binding protein. Preferred polypeptide fragments comprise the following amino acid sequence:

35 MSPRYPGGPRPLRIPNQALGGVPGSQPLPSGMDPTRQQGHPNMGGPMQR
TPPRGMVPLGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNPTNAN

SIPYSSASPGNYVGPPGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPNR
PNFPMGPGSDGPMGGLGGMESHHMNGSLSGDMDSISKSPNNMSLSNQP
GTPRDDGEMGGNFLNPfqsesySPSMTMSV (SEQ ID NO:530); MSPRYPGG
PRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRMTPPRGMV
5 LGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNPTNANSIPYSSASP
GNY (SEQ ID. NO:531); LNALGGPGMPGMNMGPGGGRPWPNPTNANSIPYSS
ASPGNYVGPPGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPN (SEQ ID
NO:532); GPMGGLGGMESHHMNGSLSGDMDSISKSPNNMSLSNQPGTPR
DDGEMGGNFLNPfqsesySPSMTMSV (SEQ ID NO:533); TCEHSSEAKAFHDY
10 (SEQ ID NO:534). Also preferred are polynucleotide fragments encoding these
polypeptide fragments. (See Accession No. 1562534)

This gene is expressed primarily in placenta and to a lesser extent in the fetal heart and a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities, fetal deficiencies, and particularly of the cardiovascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification
20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to
25 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental abnormalities or fetal deficiencies, ovarian and other endometrial cancers, reproductive dysfunction, cardiovascular disorders, and pre-natal disorders.
30

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed primarily in fetal liver and to a lesser extent in the breast and testes.
35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, liver disorders (including hepatoblastomas) and reproductive disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

5 hepatic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

10 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The

15 expression in testes and breast indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine and reproductive disorders (e.g. sperm maturation, milk production, testicular and breast cancers).

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. W93595).

This gene is expressed primarily in smooth muscle and to a lesser extent in brain.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

30 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

35 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of restenosis, atherosclerosis, stroke, angina, thrombosis, wound healing and other conditions of heart disease. In addition, the expression in brain would suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

Gene shares homology with human stromalin-2. Preferred polypeptide fragments comprise the following amino acid sequence:

QAFVLLSDLLLIFSPQMIVGGRDFLRPLVFFPEATLQSELASFMDHVFIQPGDL

15 GSGA (SEQ ID NO:535); ACSYLLCNPEFTFFSRADFARSQLVDLLTDRFQQE
LEELLQVG (SEQ ID NO:536), QKQLSSLRDRMVAFCELCQSCLSDVDTEIQEJV
ST (SEQ ID NO:537); QVILPALTLYFSILWTLTHISKSDAS (SEQ ID NO:538);
STHDLTRWELYEPCCQLLQKAVDTGXVPHQV (SEQ ID NO:539). Also preferred
are polynucleotide fragments encoding these polypeptide fragments (See Accession
20 No.R65208) This gene maps to chromosome 7, and therefore, may be used as a
marker in linkage analysis for chromosome 7 (See Accession No. D52585).

This gene is expressed primarily in the brain (infant brain, adult brain, pituitary, cerebellum, hippocampus, schizophrenic hypothalamus, amygdala).

Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, developmental and neurodegenerative diseases of the brain and nervous
system. Similarly, polypeptides and antibodies directed to these polypeptides are useful
in providing immunological probes for differential identification of the tissue(s) or cell
30 type(s). For a number of disorders of the above tissues or cells, particularly of the
central nervous system, expression of this gene at significantly higher or lower levels
may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or
bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another
tissue or cell sample taken from an individual having such a disorder, relative to the
35 standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 306 as residues: Thr-25 to Lys-36, Lys-55 to Ser-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of developmental, 5 degenerative and behavioral conditions of the brain and nervous system (e.g., schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

10 This gene is expressed primarily in the hypothalamus of a human suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 15 not limited to, disorders of the CNS particularly schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, such as schizophrenia expression of this gene at significantly higher or lower levels may be routinely detected 20 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID 25 NO: 307 as residues: Gly-38 to Ala-44.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of schizophrenia and other disorders involving the CNS.

30 **FEATURES OF PROTEIN ENCODED BY GENE NO: 75**

Preferred polypeptides of the invention comprise the following amino acid sequence encoded by this gene:

LAVSTSFICCADCISTALPLGSSRPAPAPRHRREHEHGHQARPPRLXTSLMPLSTP
AAAQLLWTQLTPMGGGRPGRHSPTLHTGPRALPPGPPHPSLHVAAALSLLR

35 (SEQ ID NO:540). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in endometrial tumor and to a lesser extent in amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and immune disorders particularly cancers of those systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 308 as residues: Ser-3 to Arg-9.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune and reproductive disorders particularly cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in kidney cortex and to a lesser extent in early stage human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders such as renal cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 309 as residues: Gly-38 to Gly-45, Gly-47 to Gly-52, Pro-92 to Lys-110.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of renal diseases such as cancer of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in kidney medulla.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of metabolic and renal diseases and disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed in chronic synovitis and microvascular endothelium.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, arthritis and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular and skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of arthritic and other inflammatory diseases as well as cardiovascular diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed in resting T-cells and activated monocytes.

- Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
10 tissues or cells, particularly of the immune system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for the study and treatment of immune diseases such as inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

- 20 This gene is expressed in a variety of immune system tissues, e.g., neutrophils,
T-cells, and TNF induced epithelial and endothelial cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, infectious and immune disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the immune and vascular systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 313 as residues: Met-1 to Trp-6.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of infectious diseases, immune and vascular disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed in activated neutrophils.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and other immune conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

- 20 This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and other immune conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 315 as residues: Ala-83 to Thr-91.

- 35 The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 10 of the above tissues or cells, particularly of the immune and inflammatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
- 15 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the inflammatory and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory and immune systems, expression of this gene at significantly higher or lower levels may be
- 25 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 30
- 35

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the immune and inflammatory systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

This gene is expressed in activated neutrophils.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and immune system diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and inflammatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of diseases of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene is expressed in activated neutrophils.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 319 as residues: Met-1 to Gly-6, Gly-32 to Pro-43, Leu-55 to Gln-60.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of disorders of the immune and inflammatory system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

In specific embodiments, polypeptides of the invention comprise the sequence:

EQVLALLWPRFELILEMNVQSVRSTDPMQLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLME
5 RAADDSSKEVESFQQLLNARTQEFIEELLSPPFGGLVAFVKEAEALIERGQAERLR
GEEARVTQLIRGFGSSWKSSVESLSQDVMRSGFTNFRNGTSIIQG (SEQ ID
NO:541), ALLKYRFFYQFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMK
VQYEEVAEKDDLMGVEDTAKKGFXSKPSRSRNTIFTLGTRGSVISPTLEAPILV
10 PHTAQR (SEQ ID NO: 542); EQRYPFEALFRSQHYXLLDNSCREYLFIGEFFFVVS
GPXAHDLFHAVMGRTLSMTLKHLDSYLAADCYDAIAVFLCIHIVLFRNIAAKRD
VPALDRYW (SEQ ID NO:543), GGLDTRPHYITRRYAEFSSALVSINQ (SEQ ID
NO:544); SRKEQLVFLINNYDMMLGVL (SEQ ID NO: 545) and/or ALLKYRFFY
QFLLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMKVQYEEVAEKDDLMG
15 VEDTAKKGFXSKPSLRSRNTIFTLGTRGSVISPTLEAPIVPHTAQRXEQRYPF
EALFRSQHYXLLDNSCREYLFIGEFFFVSGPXAHDLFHAVMGRTLSMTLKHL
DSYLAADCYDAIAVFLCIHIVLFRNIAAKRDVPALDRYWEQVLALLWPRFELILEM
NVQSVRSTDPMQLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLMER
AADDSSKEVESFQQLLN
20 ARTQEFIEELLSPPFGGLVAFVKEAEALIERGQAERLRGEEARVTQLIRGFGSSW
KSSVESLSQDVMRSGFTNFRNGTS (SEQ ID NO:546). Polynucleotides encoding
these polypeptides are also encompassed by the invention. The translation product of
this gene shares sequence homology with suppressor of actin mutation which is thought
to be important in mutation suppression.

25 This gene is expressed primarily in fetal liver and to a lesser extent in a variety
of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
30 not limited to, liver and mutations. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the liver or cancer, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
35 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 320 as residues: Val-53 to Arg-60, Thr-88 to Thr-94, Ala-142 to Ser-150, Gly-188 to Glu-196, Gly-208 to Ser-214, Thr-227 to Gly-232, Lys-279 to Phe-285.

5 The tissue distribution and homology to suppressor of actin mutation suggest that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and of liver disorder or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 88

10 This gene maps to chromosome 9, and therefore can be used in linkage analysis as a marker for chromosome 9. In specific embodiments, polypeptides of the invention comprise the sequence:
YEGKEFDYVFSIDVNEGGPSYKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVA
KFIIDNTKGQMLGLGNPSFSDPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYV
15 PGSASMGTTMAGVDPFTGNSAYRSAASKTMNITYFPKKEAVTFDQANPTQILGK
LKELNGTAPEEKKLTEDDLILLEKILSLICNSSEKPTVQQLQILWKAINCPE DIV
FPALDILRLSIKHPSVNEFCNEKEGAQFSSHLLNPKGKPANQLLALRTFC
NCFVGQAGQKLMMSQRESLMSHAIELKSGSNKNI (SEQ ID NO: 547);
HIALATLALNYSVCFHKD (SEQ ID NO: 548); HNIEGKAQCLSLISTILEVVQ
20 DLEATFRLLVALGTLISDDSNAVQLAKS (SEQ ID NO:549); LGVDSQIKKYSS
VSEPAKVSECCRFLNLL (SEQ ID NO:550); and/or YEGKEFDYVFSIDVNEGGPS
YKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVAKFIIDNTKGQMLGLGNPSFS
DPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYVPGSASMGTTMAGVDPFTGN
SAYRSAASKTMNITYFPKKEAVTFDQANPTQILGKLKELN GTAPEEKKLTEDDLI
25 LLEKILSLICNSSEKPTVQQLQILWKAINCPE DIVFPALDILRLSIKHPSVNEFC
NEKEGAQFSSHLLNPKGKPANQLLALRTFCNCFVGQAGQKLMMSQRESL
MSHAIELKSGSNKNIHIALATLALNYSVCFHKDHNIEGKAQCLSLISTILEVVQD
LEATFRLLVALGTLISDDSNAVQLAKSLGVDSQIKKYSSVSEPAKVSECCRFLN
30 LL (SEQ ID NO:551). Polynucleotides encoding these polypeptides are also
encompassed by the invention. These polypeptides share significant homology with
phospholipase A2 activating protein which is thought to be important in signal
transduction (see, e.g., Wang et al., Gene 161(2):237-241 (1995)).

35 This gene is expressed primarily in endothelial cells, to a less extent in placenta, endometrial stromal cells, osteosarcoma, testis tumor, muscle, and infant brain that are likely to be rich in blood vessels.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in vascular system, aberrant angiogenesis, tumor angiogenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system or tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in endothelial cells and several potential highly vascularized tissues and its homology to phospholipase A2 activating protein suggest that this gene may be involved in transducing signals for endothelial cells in angiogenesis or vasculogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

In specific embodiments, polypeptides of the invention comprise the sequence: YPNQDG DILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTIS
20 AYKTPRDKVQCILRCMSTIMNLLSLANEDSVPGADDFVPVLVFVLIKANPPCLL
STVQYISSFYASCLSGEESYWWMQFTAAVE (SEQ ID NO:552); YPNQDG DILR
DQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTISAYKTPRDKVQ
CILRCMSTIMNLLSLANEDSVPGADDFVPVLVFVLIKANPPCLLSTVQYISSFYA
SCLSGEESYWWMQFTAAVEFIKTI (SEQ ID NO:553); YPNQDG DILRDQVL (SEQ
25 ID NO:554); EAPWPSAQSEI (SEQ ID NO:555); PVLVFVLIKANP (SEQ ID
NO:560); SGEESYWWMQFTAAVEFIKTI (SEQ ID NO:556); ADDFVPVLVF
VLIKANPP (SEQ ID NO:557); YKTPRDKVQCIL (SEQ ID NO:558); and/or
GADDFVPVLVFVLIK (SEQ ID NO:559). The translation product of this gene shares
sequence homology with human ras inhibitor and yeast VPS9p which is thought to be
30 important in golgi vacuole transport.

This gene is expressed primarily in T cells and melanocytes and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dysfunction and disorders involving T cells and melanocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution and homology to ras inhibitor indicates that
- 10 polynucleotides and polypeptides corresponding to this gene are useful for regulating signal transduction; diagnosis and treatment of disorders involving T cells and melanocytes.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

This gene maps to chromosome 9 and therefore polypeptides of the invention can be used in linkage analysis as a marker for chromosome 9. The translation product of this gene shares sequence homology with neuronal olfactomedin-related ER localized protein which is thought to be important in influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. In specific embodiments, polypeptides of the invention comprise the sequence:

SARASTQPPAGQHPGPC (SEQ ID NO:561); MPGRWRWQRDMHPARKLLSLL
FLILMGTELTQD (SEQ ID NO:562); SAAPDSLLRSSKGSTRGSL (SEQ ID NO:563); AAIIVIWRGKSESRIAKTPGI (SEQ ID NO:564); FRGGGTLVLPPHTH
PEWLIL (SEQ ID NO:567); PLGITLPLGAPETGGGD (SEQ ID NO:565); and/or
25 CAAETWKGSQRAGQLCALLA (SEQ ID NO:566).

This gene is expressed in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and endocrinological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 323 as residues: Leu-20 to Ala-26, Arg-32 to Arg-39, Thr-104 to Gly-112.

The tissue distribution and homology to olfactomedin-related protein indicates
5 that polynucleotides and polypeptides corresponding to this gene are useful for maintenance, growth, or differentiation of neuron cells in pineal gland, therefore, may be useful for diagnosis and treatment of neurological disorders in pineal gland.

FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in prostate and apoptotic T cells.
Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate disease and T cell dysfunction. Similarly, polypeptides and
15 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detect abnormal activity in prostate and T cells or probably treatment of this abnormality.
25

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in prostate and to a lesser extent in smooth muscle cells, fibroblasts, and placenta.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in prostate or vascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
35 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating function of prostate or highly vascularized tissues, e.g. placenta.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

This gene is expressed primarily in embryos and fetal tissues stage human and to a lesser extent in a wide variety of other proliferative tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in embryonic development and cell proliferation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic tissues and proliferative cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of abnormalities in developing and proliferative cells and organs.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

The translation product of this gene shares sequence homology with transformation related protein which is thought to be important in transformation.

This gene is expressed primarily in female reproductive tissues, i.e., breast cancer cells, placenta, and ovary and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancer or dysfunction of reproductive tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction system,

5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 327 as residues: Ser-50 to Pro-61.

The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of conditions caused by transformation, i.e. tumorigenesis in

15 reproductive organs, e.g. breast, placenta, and ovary.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in testes, rhabdomyosarcoma, infant brain and to a lesser extent in some tumors and highly vascularized tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumorigenesis, abnormal angiogenesis, and/or neurological disorders. , Similarly, polypeptides and antibodies directed to these polypeptides are useful in

25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor tissues or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 328 as residues: Arg-46 to Trp-54, Pro-60 to Ile-69, Asn-116 to Ala-122, Arg-147 to Lys-153, Ser-158 to Glu-170, Ile-399 to

35 Ser-405, Pro-486 to Met-499, Pro-502 to Asp-508.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for a range of disease states including treatment of

tumor or vascular disorders and the treatment of neurological disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

This gene maps to chromosome 7 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 7. The translation product of this gene is homologous to the Clostridium perfringens enterotoxin (CPE) receptor gene product and shares sequence homology with a human 10 ORF specific to prostate and a glycoprotein specific to oligodendrocytes both of which are tissue specific proteins.(See e.g., Katahira et al., J Cell Biol. 136(6):1239-1247 (1997). PMID: 9087440; UI: 97242441.

This gene is expressed primarily in pancreas tumor and ulcerative colitis and to a lesser extent in several tumors and normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic disorder, ulcerative colitis, tumors and food poisoning. Similarly, polypeptides and antibodies directed to these polypeptides are useful in 20 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or tumorigenic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal 25 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 329 as residues: Gly-147 to Met-152, Cys-177 to Lys-188.

30 The tissue distribution and homology to prostate and oligodendrocyte-specific protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis or treatment of disorder in pancreas, ulcerative colitis, and tumors. Furthermore, identity to the human receptor for Clostridium perfringens enterotoxin indicates that the soluble portion of this receptor could be used in the 35 treatment of food poisoning associated with Clostridia perfringens by blocking the activity of perfringens enterotoxin.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

The translation product of this gene shares sequence homology with ATPase which is thought to be important in metabolism.

- 5 This gene is expressed primarily in testes and several hematopoietic cells and to a lesser extent in other tissues.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 10 not limited to, leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues
- 15 (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
- 20 NO: 330 as residues: Leu-37 to Ala-42.

The tissue distribution and homology to ATPase indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis and treatment of leukemia and other hematopoietic disorders.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 98**

In specific embodiments, polypeptides of the invention comprise the sequence: MRSARPSLGCLPSWAFSQALNI (SEQ ID NO:568); LLGLKGLAPAEISAVCE KGNFN (SEQ ID NO:569); VAHGLAWSYYIGYLRLILPELQARIR (SEQ ID NO:570); TYNQHYNNLLRGAVSQRC (SEQ ID NO:571); ILLPLDCGVPDNLSM 30 ADPNIRFLDKLPQQTGDRAGIKDRVYSN (SEQ ID NO:572); SIYELLENGQRAGT CVLEYATPLQTLFAMSQYSQAGFSGEDRLEQ (SEQ ID NO:573); AKLFCRTLE DILADAPESQNNCRLIAYQEPAADDSSFSLSQEVLRHLRQEEKEEVTVGSLKTSAV PSTSTMSQEPELLISGMEKPLPLRTDFS (SEQ ID NO:574); and/or LLGLKGLA PAEISAVCEKGNFNVAHGLAWSYYIGYLRLILPEL (SEQ ID NO:575).

35 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in prostate BPH and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, benign prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urinary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 331 as residues: Ile-60 to Asn-69, Leu-106 to Asp-112, Glu-130 to Gly-136, Phe-160 to Glu-167, Pro-184 to Cys-190, Glu-197 to Ser-202, Arg-215 to Glu-221, Thr-237 to Pro-242.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of benign prostatic hypertrophy or prostate cancer.

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders or injuries of the salivary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders of, or injuries to the salivary gland or other glandular tissue.

FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene maps to chromosome 15, accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 15. The translation product of this gene shares sequence homology with a *C.elegans* gene of unknown function. In specific embodiments, polypeptides of the invention comprise the sequence: DPRVRLNSLTCKHIFISLTQ (SEQ ID NO:583); TMKLLKLRRNIV KLSLYRHFTN (SEQ ID NO:576); TLILAVAASIVFIIWTTMKFRI (SEQ ID NO:577); VTCQSDWRELWVDDAIWRLLFSMILFVI (SEQ ID NO:578); MVLWR PSANNQRFAFSPLSEEEEDEQ (SEQ ID NO:580); KEPMLKESFEGMKMRS TKQEPNGNSKVNKAQEDDL (SEQ ID NO:584); and/or KWVEENVPSSVTDVALP ALLDSDEERMITHFERSKME (SEQ ID NO:582). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thyroid and to a lesser extent in osteoclastoma, kidney medulla, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thyroid dysfunction or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 333 as residues: Lys-107 to Leu-124, Glu-150 to Thr-159, Pro-173 to Asp-179, Ser-192 to Ser-201.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of thyroid dysfunction or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 101

This gene maps to chromosome 16, therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 16. In specific embodiments, polypeptides of the invention comprise the sequence:

IRHELTVLRDTRPACA (SEQ ID NO:585); and/or MDFXMALIYD (SEQ ID NO:586). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney cortex and to a lesser extent in adult
5 brain, corpus colosum, hippocampus, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to
10 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neurological
20 disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 102

In specific embodiments, polypeptides of the invention comprise the sequence:
MQEMMRNQDRALSNLESIPGGYNA (SEQ ID NO:587); LRRMYTDIQEPMLSA

25 AQEQF GGNPF (SEQ ID NO:588); ASLVSNTSSGEGSQPSRTENRDPLPNPWAP
QT (SEQ ID NO:589); SQSSSASSGTASTVGGTTGSTASGTSGQSTTAPNLVPGV
GASMFnTPG MQSLLQQITENPQLMQNMLSAPY (SEQ ID NO:590);
MRSMMQSLSQNPDLAAQMMLNNPLFAGNPQLQEQRQLPTFLQQ (SEQ ID
NO:591); MQNPDTLSAMSNPRAMQALLQIQQQGLQTLATEAPGLIPGFTPGLG
30 ALGSTGGSSGTNGSNATPSENTSPTAGT (SEQ ID NO:592); TEPGHQQFI
QQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALIATGGDINAA
IERLLGSQPS (SEQ ID NO:593); RNPAMMQEMMRNQDRALSNLESIPGGY
NALRRMYTDIQEPMLSAA (SEQ ID NO:594); GNPFASLVSNTSS (SEQ ID
NO:595); ENRDPLPNPWA (SEQ ID NO:595); GKILKDQDTLSQHGIHD (SEQ ID
35 NO:597); GLTVHLVIKTQNRP (SEQ ID NO:598); SELQSQMQRQLLSNPEMM
(SEQ ID NO:599); PEISHMLNNPDIMR (SEQ ID NO:600); and/or
RQLIMANPQMQLIQRNP (SEQ ID NO:601). Polynucleotides encoding these

polypeptides are also encompassed by the invention.

This gene is expressed primarily in breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 103

The translation product of this gene shares sequence homology with secreted serine proteases and lysozyme C precursor, which is thought to be important in bacteriolytic function. In specific embodiments, polypeptides of the invention comprise the sequence: NLCHVDCQDLLNPNLLAGIHCARKRIVS (SEQ ID NO:602); LDGFEGYSLSDWLCLAFVESKFN (SEQ ID NO:603); NENADGSFDYGLFQINSHYWCN (SEQ ID NO:604); and/or NLCHVDCQDLLNPNLLAGIHCARKRIVS (SEQ ID NO:605). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Ile-62 to Phe-70, Asn-
5 78 to Asn-84.

The tissue distribution and homology to lysozyme C precursor indicates that polynucleotides and polypeptides corresponding to this gene are useful for boosting the monocyte-macrophage system and enhance the activity of immunoagents.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 104

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
15 not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
20 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some immune disorders.
25

FEATURES OF PROTEIN ENCODED BY GENE NO: 105

The translation product of this gene shares sequence homology with ARI
30 protein of Drosophila (accession 2058299; EMBL: locus DMARIADNE, accession X98309), which is thought to be important in axonal path-finding in the central nervous system. In specific embodiments, polypeptides of the invention comprise the sequence IREVNEVIQNPAT (SEQ ID NO:606); ITRILLSHFNWDKEKLMERYF
35 DGNLEKLFA (SEQ ID NO:607); NTRSSAQDMPCQICYLNYPNSYF (SEQ ID NO:608); TGLECGHKFCMQCWSEYLTTKIMEEGMGQTISCPAHG (SEQ ID NO:614); CDILVDDNTVMRLITDSKVKLKYQHLITNSFVECNRLLKWCAPD CHHVVKVQYPDAKPV (SEQ ID NO:609); CDILVDDNTVMRLITDSK

- VKLKYQHLITNSFECNRLLKWCAPDCHHVVKV (SEQ ID NO:610);
GCNHMVCRNQNCKAEFCWVCLGPWEPHGSAWYNCRYNEDDAKAARDAAQE
RSRAALQRYL (SEQ ID NO:611); FYCNRYMNHMQSLRFEHKL YAQVKQ
KMEEMQQHNMSWIEVQFLKKAVDVLQCRCATLMT (SEQ ID NO: 612);
5 YVFAFYLKKNNQSIIFENNQADLENATEVLSGYLERDISQDSLQDIKQKVQDKY
RYCESR (SEQ ID NO:613) Polynucleotides encoding these polypeptides are also
encompassed by the invention.

This gene is expressed primarily in adult brain, and to a lesser extent in endometrial tumor, melanocytes, and infant brain.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases or injuries involving axonal path development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
15 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
20 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution and homology to ARI protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of
25 disease states or injuries involving axonal path development, including neurodegenerative diseases and nerve injury.

FEATURES OF PROTEIN ENCODED BY GENE NO: 106

- The translation product of this gene shares sequence homology with cytochrome
30 b561 [Sus scrofa] which is thought to be an integral membrane protein of neuroendocrine storage vesicles of neurotransmitters and peptide hormones.

This gene is expressed primarily in frontal cortex and to a lesser extent in rhabdomyosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 339 as residues:

10 Ser-18 to Pro-24.

The tissue distribution and homology to cytochrome b561 [Sus scrofa] indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of neurological disorders. This gene may also be important in regulation of some types of cancers.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 107

In specific embodiments, polypeptides of the invention comprise the sequence:

MWGYLFVDAAWNFLGCLICGW (SEQ ID NO:615); MHFISSGNVSAIRSSILL
RXSLSYLGNCRLVSAIFVYFLLFLLS (SEQ ID NO:616); and/or MDQALRGSPSE
20 GFSTDPSPPQVGRQIPSFPPWRRLVLPKASGCFLEREWWLCVFKLRTRPGAEA
HAYNSSLGGRGKGIT (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pancreas tumor and to a lesser extent in cerebellum.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

30 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

35 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 340 as residues: Pro-22 to Phe-33.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pancreatic tumors.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene maps to chromosome 17 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 10 MLPALASCCHFSPPEQAARLKKLQEQQKVEFRKRMEKEVSDFIQDSGQIK
KKFQPMNKIERSILHDVVEAGLTSFSFGEDDDCRYVMIFKKEFAPSDEELDSY
RRGEEWDPQKAEEKRNXKELAQRQ (SEQ ID NO:618); EEEAAQQGPVVV
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
IRAKKRLRQSGE (SEQ ID NO:619); PPRRPAQLPLTPGAGQGAGRDKAAAIRA
15 HPGAPPLNHLLP (SEQ ID NO:620); AVPQAGGKQVFDSLSPLELGYVRGMCVCV
(SEQ ID NO:621) and/or MLPALASCCHFSPPEQAARLKKLQEQQKVEFRK
RMEKEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEAGLTSFSFGEDDDCRYV
MIFKKEFAPSDEELDSYRRGEWDPQKAEEKRNXKELAQRQEEEAAQQGPVVV
SPASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
20 IRAKKRLRQSGE (SEQ ID NO:622). Polynucleotides encoding these polypeptides are also encompassed by the invention. The translation product of this gene shares sequence homology with FSA-1 which may play a role as a structural protein component of the acrosome.

This gene is expressed primarily in fetal kidney and sperm.

- 25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders, especially involving acrosomal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
30 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
35 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an

individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 341 as residues: Glu-8 to Asn-35.

The tissue distribution and homology to FSA-1 indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of infertility due to acrosomal dysfunction of sperm.

FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in pituitary and to a lesser extent in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 342 as residues: Met-1 to Trp-6.

Because the gene is found in both pituitary and epididymus, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of male reproductive disorders. This may involve a secreted peptide produced in the pituitary targeting the epididymus.

FEATURES OF PROTEIN ENCODED BY GENE NO: 110

In specific embodiments, polypeptides of the invention comprise the sequence: LLCPVNLNSGXSWNFPHPSQPEYSFHGFHSTRLWI (SEQ ID NO:623); and/or PSTPWFLFLLGLTCPFSTSHPRWDSIPP (SEQ ID NO:624). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in resting T-cells. .

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

- not limited to, T-cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or
- 5 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of certain immune disorders, especially those involving T-cells.
- 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 111**
- This gene is expressed primarily in cerebellum and whole brain and to a lesser extent in infant brain and fetal kidney.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 20 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
- 25 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 344 as residues:
- 30 Asp-48 to Gly-55.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 112

The translation product of this gene shares sequence homology with yeast mitochondrial ribosomal protein homologous to ribosomal protein s15 of E.coli which

is thought to be important in the early assembly of ribosomes (See Accession No. M38016). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in developmental tissues.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development of cancers and tumors in addition to healing wounds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
- 10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
- 15 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribosomal protein s15 of *E. coli* indicates that polynucleotides and polypeptides corresponding to this gene are useful for

20 diseases related to the assembly of ribosomes in the mitochondria which is important in the translation of RNA into protein. Therefore, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of multiple tumors as well as in healing wounds which are thought to be under similar regulation as developmental tissues. Protein, as well as, antibodies directed against the

25 protein have utility as tumor markers, in addition to immunotherapy targets, for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 113

The translation product of this gene shares sequence homology with human

30 poliovirus receptor precursors which are thought to be important in viral binding and uptake. Preferred polypeptide fragments comprise the following amino acid sequence: ELSISISNVALADEGEYTCISIFTMPVRTAKSLVTVLGIPQKPIITGYKSSLREKDT ATLNCQSSGSKPAARLTWRKGDQELHGEPTRIQEDPNGKTFTVSSSVTFQVTR EDDGASIVCSVNHESLKGADRSTSQRIEVLYTPTAMIRPDPPHPREGQKLLLHC

35 EGRGNPVPQQYLWEKEGSVPPLKMTQESALIFPFLNKSDSGTYGCTATSNMGS YKAYYTLNVND (SEQ ID NO:625). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gnl|PID|d1002627).

This gene is expressed almost exclusively in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, susceptibility to viral disease and diseases of the CNS especially cancers of that system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 346 as residues: Leu-26 to Asp-37, Lys-53 to Ser-59.

The tissue distribution and homology to poliovirus receptor precursors indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and prevention of diseases that involve the binding and uptake of virus particles for infection. It might also be helpful in genetic therapy where the goal is to insert foreign DNA into infected cells. With the help of this protein, the binding and uptake of this foreign DNA might be aided. In addition, it is expected that over expression of this gene will indicate abnormalities involving the CNS, particularly cancers of that system.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene shares sequence homology with YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* in addition to alpha-1 collagen type III (See Accession No. gil537432). One embodiment for this gene is the polypeptide fragment(s) comprising the following amino acid sequence: VPELPDRVHQLHQAVQGCALGRPGFPGGPTH SGHHKSHPGPAGGDYNRCDRPGQVHLHNPRGTGRRQLHPTAGPGVHRRA CPSQQLPHRLPGVPCPSLTPVLPSWTQSWCG LPGYTSSS (SEQ ID NO:630). An additional embodiment is the polynucleotide fragment(s) encoding these polypeptide fragments

This gene is expressed primarily in brain cells and to a lesser extent in activated B and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegeneration and immunological disorders. Similarly, polypeptides 5 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, 10 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 347 as residues: Glu-34 to Glu-39, Gly-51 to Ser-72, Ala-88 to Glu-93, Gln-100 15 to Val-105.

The tissue distribution and homology to YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* as well as to a conserved alpha-1 collagen type III protein indicates that polynucleotides and 20 polypeptides corresponding to this gene are useful for the detection and treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune 25 functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 115

The translation product of this gene shares sequence homology with alpha 3 type IX collagen which is thought to be important in hyaline cartilage formation via its 30 ability to uptake inorganic sulfate by cells (See Accession No. gi|975657). One embodiment of this gene is the polypeptide fragment comprising the following amino acid sequence: SLRRPRSAAXQTLTTFLSSVSSASSALPGSREPCDPRAPPPR SGSAASCCSCCSCPRRRAPLSPRGSKRRIRQREVVDLYNGMCLQGPAGVPG RDGSPGANGIPGTPGIPGRDGFKGEKGECLRESFEESWTPNYKQCSWSSLNY 35 GIDLGKIAECTFTKMRNSALRVLFSGSLRLKCRNACCQRWYFTNGAECSPGP LPIEAIYLDQGSPEMNSTINIHTSSVEGLCEGIGAGLVDAIWVGTCSDYPKG DASTGWNSVSRIIEELPK (SEQ ID NO:634). An additional embodiment are the

polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in smooth muscle and to a lesser extent in synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e., spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 3 type IX collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases associated with the mutation in this gene which leads to the many different types of chondrodysplasias. By the use of this product, the abnormal growth and development of bones of the limbs and spine could be routinely detected or treated in utero since the protein or mureins thereof could affect epithelial cells early in development and later the chondrocytes of the developing craniofacial structure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 116

The translation product of this gene shares sequence homology with retrovirus-related reverse transcriptase which is thought to be important in viral replication. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: TKKENCRRPASLMNIDTKILNKILMNQ (SEQ ID NO:640). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. pirlA25313IGNHUL1).

This gene is expressed primarily in human meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, retroviral diseases such as AIDS, and possibly certain cancers due to transactivation of latent cell division genes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to retrovirus-related reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of diseases and maladies associated with retroviral infection since a functional reverse transcriptase (RT) or RT-like molecule is an integral component of the retroviral life cycle.

FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with an unknown gene from *C. elegans*, as well as weak homolog with mammalian metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is known to be required for embryonic development. Preferred polypeptide fragments comprise the following amino acid sequence: MCNLPIKVVCRANAЕYMSPSGKVPXXHVGNQ VVSELGPIVQFVKAKGHSLSDGLEEVQKAEMKAYMELVNNMLLTAELYLQWC DEATVGXITHXRYGSPYPWPLXHILAYQKQWEVKRXXKAIGWGKKLDQVLE 25 DVDQCCQALSQRQLGTQPYFFNKQPTELDALVFGHLYTILTQLTNDELSEKVKN YSNLLAFCRRI EQHYFEDRGKGRLS (SEQ ID NO:641); MCNLPIKVVCRANAЕ YMSPSGKVPXXHVGNQVVSELGPIVQFVK (SEQ ID NO:642). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gil1326108).

This gene is expressed primarily in fetal tissues and to a lesser extent in hematopoietic cells and tissues, including spleen, monocytes, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer; lymphoproliferative disorders; inflammation; chondrosarcoma, and Gaucher disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and embryonic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 118

The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA chain from an RNA molecule, and is a method whereby the infecting RNA chains of retroviruses are transcribed into their DNA complements. One embodiment for this gene is the polypeptide fragment comprising the following amino acid sequence:
MXXXNSHITIFTLNVNGLNAPNERHRLANWIQSQDVCCIQETHLTGRDTHRL
KIKGWRKTYQANGKQKK (SEQ ID NO:647). An additional embodiment is the polynucleotide fragments comprising polynucleotides encoding these polypeptide fragments (See Accession No. gil2072964).

This gene is expressed primarily in skin and to a lesser extent in neutrophils. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, hematopoietic disorders; inflammation; disorders of immune surveillance. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epidermis and/or hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and

wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 The tissue distribution and homology to reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for cancer therapy. Expression in the skin also indicates that this gene is useful in wound healing and fibrosis. Expression by neutrophils also indicates that this gene product plays a role in inflammation and the control of immune surveillance (i.e. recognition of viral
10 pathogens). Reverse transcriptase family members are also useful in the detection and treatment of AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 119

- The translation product of this gene shares sequence homology with reverse
15 transcriptase which is important in the synthesis of a cDNA copy of an RNA molecule, and is a method whereby a retrovirus reverse-transcribes its genome into an inheritable DNA copy.

- This gene is expressed primarily in the frontal cortex of brain.
Therefore, polynucleotides and polypeptides of the invention are useful as
20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
25 of the above tissues or cells, particularly of the CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to reverse transcriptase suggest that this is useful in the treatment of cancer and AIDS. The expression in brain indicates that it plays a role in neurodegenerative disorders and in neural degeneration.

FEATURES OF PROTEIN ENCODED BY GENE NO: 120

One embodiment of this gene has homology to a hypothetical protein in *Schizosaccharomyces pombe* (See Accession No. 2281980). Another embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

5 IYHLHSWIFFHFKRAFCMCFITMKVIHAHCSKLRKCXNAQISVFCTTLTASYPT
(SEQ ID NO:651). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

10 This gene is expressed primarily in adult hypothalamus and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disorders; endocrine function; and vertigo. Similarly, 15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded 20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of neurodegenerative disorders; diagnosis of tumors of a brain or neuronal origin; treatments involving hormonal control of the entire body and of homeostasis, behavioral disorders, such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and 30 panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 121

35 The translation product of this gene shares sequence homology with the human IRLB protein which is thought to be important in binding to a c-myc promoter element and thus regulating its transcription (See Accession No. gil33969). This gene maps to

chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in brain and breast and to a lesser extent in a variety of hematopoietic tissues and cells.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer of the brain and breast; lymphoproliferative disorders; neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these
- 10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, breast, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
- 15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancer of the
- 20 brain, breast, and hematopoietic system. In addition, it may be useful for the treatment of neurodegenerative disorders, as well as disorders of the hematopoietic system, including defects in immune competency and inflammation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with an ATP synthase, a key component of the proton channel that is thought to be important in the translocation of protons across the membrane.

- 30 This gene is expressed primarily in T cell lymphoma. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, T cell lymphoma. Similarly, polypeptides and antibodies directed to these
- 35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATP synthase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of defects in proton transport, homeostasis, and metabolism, as well as the diagnosis and treatment of lymphoma. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene maps to chromosome 15, and therefore, may be used as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in a variety of fetal tissues, including fetal liver, lung, and spleen, and to a lesser extent in a variety of blood cells, including eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer (abnormal cell proliferation); T cell lymphomas; and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetus and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions involving cell proliferation. Expression of this gene in fetal tissues, as well as in a variety of blood cell lineages indicates that it may play a role in either cellular proliferation; apoptosis; or cell survival. Thus it may be useful in the management and

treatment of a variety of cancers and malignancies. In addition, its expression in blood cells suggest that it may play additional roles in hematopoietic disorders and conditions, and could be useful in treating diseases involving autoimmunity, immune modulation, immune surveillance, and inflammation..

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in placenta and to a lesser extent in pineal gland and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, endocrine, and female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 357 as residues: Leu-69 to Val-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders in development. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in benign prostatic hyperplasia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of benign prostatic hyperplasia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive

- system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene
- 5 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of benign prostatic hyperplasia. Protein, as well as, antibodies directed against the protein may show utility

10 as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in apoptotic T-cells and to a lesser extent in suppressor T cells and ulcerative colitis.

- 15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving premature apoptosis, and immunological and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these
- 20 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
- 25 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders involving inappropriate levels of apoptosis, especially in immune cell lineages. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in Raji cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and T cell autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 360 as residues: Asp-23 to Gly-29.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammation and T cell autoimmune disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 128

- The translation product of this gene shares sequence homology with an *C. elegans* coding region C47D12.2 of unknown function (See Accession No. gnl|PID|e348986). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: EDDGFNRSIHEVILKNITWY SERVLTEISLGSLLILVVIRTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQY 30 AAQRIISLFSLLSKKHNVLEQATQSLRGSLSSNDVPLPDYAQDLNVIEEVIRMM LEIINSCLNSLHHNPNLVALLYKRDLFEQFRTHPSFQDIMQNIDLVISFFSSRLL QAGS (SEQ ID NO:657); EDDGFNRSIHEVILKNITWYSERVLTEISLGSLLILVV (SEQ ID NO:658); RTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQYAAQ RIISLFSLLSKKH (SEQ ID NO:659); KKHNKVLEQATQSLRGSLSSNDVPLPDY 35 AQD (SEQ ID NO:661); SCLTNSLHHNPNLVYALLYKRDLFEQFRTHPSFQD IMQNIDLVISFFSSRLLQAGS (SEQ ID NO:660). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to

chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in smooth muscle and to a lesser extent in fetal liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, atherosclerosis and other cardiovascular and hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of circulatory system
20 disorders such as atherosclerosis, hypertension, and thrombosis . In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the
25 expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 129

30 The translation product of this gene shares sequence homology with a ribosomal protein which is thought to be important in cellular metabolism, in addition to the *C.elegans* protein F40F11.1 which does not have a known function at the current time (See Accession No. gnl|PIDle244552). Preferred polypeptide fragments comprise the following amino acid sequence:
35 MADIQTERAYQKQPTIFQNKKRVLLGETGKEKLPRVTNKNIGLGFKDT
PRRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQDEDAEDHCHPPRLSALHPQVQ
PLREAPQEHVCTPVPL LQGRPDR (SEQ ID NO:662); MKMQRTIVIRRDLH

YIRKYNRFEKRHKNMSVHLSPCFRDVQIGDIVTVGECRPLSKTVRFNVLKVTK
AAGTKKQFQKF (SEQ ID NO:663); MADIQTERAYQKQPTIFQNKKRVLLGET

GK (SEQ ID NO:664); HCHPPRLSALHPQVQPLREAPQEHVCTPVPL LQGRPDR

(SEQ ID NO:666); NIGLGFKDTPRRLRGTYIDKKCPFTGNVSIRGRILSGVVTQ

5 (SEQ ID NO:669); MKMQRTIVIRRDLHYIRKYNRFEKRHKNMSVHLSP (SEQ
ID NO:667); CFRDVQIGDIVTVGECRPLSKTVRFNVLKVTKAAGTKKQFQKF
(SEQ ID NO:668). Also preferred are polynucleotide fragments encoding these
polypeptide fragments.

This gene is expressed primarily in Wilm's tumor and to a lesser extent in
10 thymus and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, diseases affecting RNA translation. Similarly, polypeptides and

15 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the Wilm's tumors, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 362 as residues:

Thr-11 to Asp-20.

25 The tissue distribution and homology to a ribosomal protein indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diseases
affecting RNA translation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 130

30 The translation product of this gene shares sequence homology with a yeast
DNA helicase which is thought to be important in global transcriptional regulation (See
Accession No. gnl|PIDle243594). One embodiment for this gene is the polypeptide
fragments comprising the following amino acid sequence: IFYDSDWNPTVDQQA
MDRAHRLGQTQVTVYRICKGTIEERILQRAKEKSEIQRMVISG (SEQ ID
35 NO:670); TRMIDLLEEYMYRKHTYXRLDGSSKISERRDMVADFQNRNDI
FVFLLSTRAGGLGINLTAXDTVHF (SEQ ID NO:671); TRMIDLLEEYMYRK
HTYXRLDGSSKISERRDM (SEQ ID NO:674); RRDMVADFQNRNDIFVFL

STRAGGLGINLTAXDTVHF (SEQ ID NO:675) , IFYDSDWNPTVDQQAMD
RAHRLGQTKQVTVYRLICKG (SEQ ID NO:676); RLICKGTIEERILQRAK
EKSEIQRMVISG (SEQ ID NO:678). An additional embodiment is the polynucleotide
fragments encoding these polypeptide fragments.

5 This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases and disorders of the brain. Similarly, polypeptides and 10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to a DNA helicase indicates that 20 polynucleotides and polypeptides corresponding to this gene are useful for diseases affecting RNA transcription, particularly developmental disorders and healing wounds since the later are thought to approximate developmental transcriptional regulation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 131

25 This gene is expressed primarily in prostate and to a lesser extent in amygdala and pancreatic tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 30 not limited to, prostate enlargement and gastrointestinal disorders, particularly of the pancreas and gall bladder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or 35 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of prostate diseases, 5 including benign prostatic hyperplasia and prostate cancer. In addition, the tissue distribution in tumors of the pancreas indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or 10 immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in adult lung and to a lesser extent in hypothalamus.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary diseases and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological 20 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample 25 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pulmonary and 30 respiratory disorders such as emphysema, pneumonia, and pulmonary edema and emboli. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, 35 obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 133

- 5 This gene is expressed primarily in human liver. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cirrhosis of the liver and other hepatic disorders. Similarly, polypeptides 10 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides 20 corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 134**

This gene is expressed primarily in fetal kidney and to a lesser extent in fetal liver and spleen.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development and regeneration of liver and kidney and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and excretory systems, expression of this gene at significantly higher or 35 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 367 as residues: Pro-70 to Arg-77, Tyr-
5 102 to Thr-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the kidney and liver, such as cirrhosis, kidney failure, kidney stones, and liver failure, hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are
10 attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in brain, bone marrow, and to a lesser extent in placenta, T cell, testis and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and immunological diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and
25 immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 368 as residues: Met-1 to His-6.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
35 Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also

play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 136

- 5 Translatation product of this gene is homologous to the human WD repeat protein HAN11. Preferred polypeptide fragments comprise the following amino acid sequence:
- MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFVEEYNNKVQLVG
LDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDYLRVWRVGETET
10 RLECLLNNNKNSDFCPLTSFDWNEVDPYLLGTSSIDTTCTIWGLETQVLGRV
NLVSGHVKTQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEH
STIIYEDPQHHPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTIE
HVSMALLGPHIHPATSALQRMTTRLSGTSSKCPEPLRTLSWPTQLXGEINNVQ
WASTQPELSPSATTAWRYSECSVGGAVPTRQGLLYFLPLPHPQS (SEQ ID
15 NO:679); MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFV
EEYNNKVQLVGLDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDY
LRVWRVGETETRLECLLNNNKNSDFCPLTSFDWNEVDPYLL (SEQ ID
NO:680); SFDWNEVDPYLLGTSSIDTTCTIWGLETQVLGRVNLVSGHV
TQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEHSTIIYEDPQH
20 HPPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTI (SEQ ID
NO:681); VGADGSVRMFDLRHLEHSTIIYEDPQHHPLLRLCWNKQDPNYLA
TMAMDGMEVVILDVRVPAHLXPGTTIEHVSMALLGPHIHPATSALQRMTTRLS
SGTSSKCPEPLRTLSWPTQLXGEINNVQWASTQPELSPSATTAWRYSECSVG
GA VPTRQGLLYFLPLPHPQS (SEQ ID NO:682). Also preferred are polynucleotide
25 fragments encoding these polypeptide fragments.

This gene is expressed primarily in placenta, embryo, T cell and fetal lung and to a lesser extent in endothelial, tonsil and bone marrow.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and developmental diseases in addition to cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 369 as residues: Gly-19 to Gln-28, Pro-36 to Phe-42.

- 5 The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above
10 listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

This gene is expressed primarily in TNF and INF induced epithelial cells, T cells and kidney.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory conditions particularly inflammatory reactions in the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of renal
25 system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 370 as residues: Thr-67 to Gly-72, Gln-132 to Ala-145, Arg-150 to Pro-157.

The tissue distribution indicates that the protein products of this gene are useful for treating the damage caused by inflammation of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. D63485).

- 5 This gene is expressed primarily in breast cancer and colon cancer and to a lesser extent in thymus and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers, especially of the breast and colon tissues. Similarly,

- 10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 15

- 20 The tissue distribution in tumors of colon and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 139**

This gene maps to chromosome 17, and therefore, can be used as a marker for linkage analysis from chromosome 17.

This gene is expressed primarily in CD34 positive cells, and to lesser extent in activated T-cells and neutrophils.

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunologically related diseases and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoietic system, expression of this gene at significantly higher or lower levels
- 35

may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

5 fluid from an individual not having the disorder.

The tissue distribution in CD34, T-cell and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of hematopoietic disorders and immunologically related diseases, such as anemia, leukemia, inflammation, infection, allergy, immunodeficiency disorders,

10 arthritis, asthma, immune deficiency diseases such as AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene was recently cloned by another group, who called the gene KIAA0313 gene. (See Accession No. d1021609.) Preferred polypeptide fragments comprise the amino acid sequence:

LYATATVISSPSTEXLSQDQGDRASLDAADSGRGSWTSCSSGSHDNIQTIQ
HQRSWETLPFGHTHFDYSGDPAGLWASSSHMDQIMFSDHSTKYNRQNQSRES
LEQAQSRASWASSTGYWGEGDSEGDTGTIKRRGGKDVSIEAESSSLTSVTTEETK
PVPMMPAHIAVASSTTKGLIARKEGRYREPPPPTPPGYIGIPITDFPEGHSHPARKP
20 PDYNVALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQWHKXNESDPR
LAPYQSQGFSTEEDDEEQVSAV (SEQ ID NO:683); HMDQIMFSDHSTKYNRQ
NQSRESLEQAQSRASWASSTGYWGE (SEQ ID NO:684); SVTTEETKPVPMP
AHIAVASSTTKGLIARKEGRYREPPPPTPPGYIGIPITD (SEQ ID NO:685); and
VALQRSRMVARSSDTAGPSSVQQPHGHPTSSRPVNKPQW
25 HKXNESDPR LAPYQSQGF (SEQ ID NO:686). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 4, and therefore, may be used as a marker in linkage analysis for chromosome 4 (See Accession No. AB002311).

This gene is expressed primarily in ovarian cancer, tumors of the Testis, brain, and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, ovarian, testicle, brain and colon cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male and female reproductive systems,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

- 5 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, testis, and brain origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where
10 expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 141

- 15 This gene is expressed primarily in spleen and colon cancer. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, colon cancer and immunological disorders. Similarly, polypeptides and
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal tract and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
25 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution in tumors of colon, ovary, and breast origins indicates
30 that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 142

Translation product is homologous to T cell translocation protein, a putative zinc finger factor (See Accession No. 340454), as well as to the G-protein coupled receptor TM5 consensus polypeptide (See Accession No. R50734). Preferred polypeptide

5 fragments comprise the following amino acid sequence:

CLLFVFVSLGMRCLFWTIVNVLYLKHKCNTVLLCYHLCI (SEQ ID NO:687); ACSKLIPAFEMVMRAKDNVYHLDCAFQLCNQRXCVGDKFFLKNNXLCQT DYEEGLMKEGYAPXVR (SEQ ID NO:688). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

10 This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders including brain cancer. Similarly, polypeptides 15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central Nervous System, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, 20 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides 25 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with 30 the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 143

Translation product for this gene has significant homology to the Fas ligand, which is a cysteine-rich type II transmembrane protein/tumor necrosis factor receptor 35 homolog. Mutations within this protein have been shown to result in generalized lymphoproliferative disease leading to the development of lymphadenopathy and autoimmune disease (See Medline Article No. 94185175). Preferred polypeptide

fragments comprise the following amino acid sequence:

SALSEPGAPDRRRPCPESVPRPDDEQWPPPPTALCLDVAPLPPSS (SEQ ID NO:689). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. 473565).

5 This gene is expressed primarily in osteoblasts, lung, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoblast-related, pulmonary, neurological, and immunological

10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded 15 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 376 as residues: Trp-33 to Thr-40, Lys-20 45 to Ile-63.

The tissue distribution in osteoblasts, lung, and brain combined with its homology to the Fas ligand indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, 25 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the Fas ligand gene is known to be expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including asthma, immune deficiency diseases such as AIDS 30 and leukemia, and various autoimmune disorders including lupus and arthritis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene shares sequence homology with a 21.5 KD transmembrane protein in the SEC15-SAP4 intergenic region of yeast. (See Accession No. 1723971.) Preferred 35 polypeptide fragments comprise the amino acid sequence:

AHASESGERWWACCGVRFGLRSIEAIGRSCCHDPGGLVANRGRRFKWAIEL
SGPGGGSRGRSDRGSGQGDSLYPVGYLDKQVPDTSVQETDRILVEKRCWDIAL

GPLKQIPMNLFIMY MAGNTISIFPTMMVCMMAWRPIQALMAISATFKMLESSSQ
KFLQGLVYLIGNLMGLALAVYKCQSMGLLPTHASDWLAFIEPPERMEFSGG
GLLL (SEQ ID NO:691); PVGYLDKQVPDTSVQETDRILVEKRCWDIALGPLKQ
IPMNLFI (SEQ ID NO:693); and ATFKMLESSSQKFLQGLVYLIGNLMGLALAV

5 YKCQSMGLLPTHASD (SEQ ID NO:692). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma, hemangiopericytoma, liver, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoclastoma, hemangiopericytoma, liver and lung tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the above tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lung and liver systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard 15 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing osteoclastoma, hemangiopericytoma, liver and lung tumors.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 145

Translation product of this gene shares homology with the glucagon-69 gene which may indicate this gene plays a role in regulating metabolism. (See Accession No. A60318) One embodiment for this gene is the polypeptide fragments comprising the 30 following amino acid sequence:

PTTKLDIMEKKKHIQIRFPSFYHKLVDSGRMRSKRETRREDSDTKHNL (SEQ ID NO:694). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain, kidney, colon, and testis.

35

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, brain, kidney, colon, and testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, neurological, circulatory, and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of brain, kidney, colon, and testis origins, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 146

The translation product of this gene shares sequence homology with goliath protein which is thought to be important in the regulation of gene expression during development. Protein may serve as a transcription factor. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

30 TEHIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGSLSFVFSISFIV
LMISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKE
PDFDHCAVCIESYKQNDVVRILPCKHVFKSCVDPWLSEHCTPMCKLNILKA
LGIV (SEQ ID NO:695); TEHIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMP
PKNFSRGSLSFVFSISFIVLM IISSAWLIFYF (SEQ ID NO:697); SISFIVLMISSA
35 WLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTVKKGDKE (SEQ ID
NO:698); VKKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFKSCVDP

WLSEHCTCPMCKLNILKALGIV (SEQ ID NO:699). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. 157535). Moreover, another embodiment is the polynucleotide fragments encoding these polypeptide fragments:

- 5 MTHPGTEHIIAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGS
LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRRLGDAAKKAISKLTTRTV
KKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHKSCVDPWLSEHCTCP
MCKLNILKALGIVPNLPCTDNVAFDMERLRTQAVNRRSALGDLAGDNSLGLE
PLRTSGISPLPQDGELTPRTGEINIAVTKEWFIASFGLLSALTLCYMIIRATASLN
10 ANEVEWF (SEQ ID NO:696); MTHPGTEHIIAVMITELRGKDILSYLEKNISVQM
TIAVGTRMPPKNFSRGLVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR
LGDAAKKAISKLTTRTVKKGDKE
TDPDFDHCAVCIESYKQNDVVRILPCKHKSCVDPWLSEHCTCPMCKLNIL
KALGIVPNLPC (SEQ ID NO:701); TQAVNRRSALGDLAGDNSLGLEPLRTSGI
15 SPLPQDGELTPRTGEINIAVTKEWFIASFGLLSALTLCYMIIRATASLNANEVEWF
F (SEQ ID NO:702); PLHGVADHLGCDPQTRFFVPPNIKQWIALLQRGNCTF
KEKISRAAFHNAVAVVITYNNKSKEEPVTMTHPGTEHIIAVMITELRGKDILSYLE
KNISVQMTIAVGTRMPPKNFSRGLVFVSISFIVLMISSAWLIFYFIQKIRYTN
RDRNQRRLGDAAKKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVR
20 LPCKHVKSCVDPWLSEHCTCPMCKLNILKALGIVPNLPCTDNVAFDMERLT
RTQAVNRRSALGDLAGDNSLGLEPLRTSGISPLPQDGELTPRTGEINIAVTKEW
FIASFGLLSALTLCYMIIRATASLNANEVEWF (SEQ ID NO:703); and
HGVADHLGCDPQTRFFVPPNIKQWIALLQRGNCTFKEKISRAAFHNAVAVVITY
NNKSKEE (SEQ ID NO:704). An additional embodiment is the polynucleotide
25 fragments encoding these polypeptide fragments. When tested against Jurkat cell lines,
supernatants removed from cells containing this gene activated the GAS pathway.
Thus, it is likely that this gene activates immune cells through the JAKS/STAT signal
transduction pathway.

This gene is expressed primarily in macrophage, breast, kidney and to a lesser
30 extent in synovium, hypothalamus and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, schizophrenia and cancer. Similarly, polypeptides and antibodies directed
35 to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune and neural system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 5 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zinc finger protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of schizophrenia, kidney disease and other cancers. The tissue distribution in macrophage, breast, and kidney origins indicates that polynucleotides and polypeptides 10 corresponding to this gene are useful for diagnosis and intervention of tumors within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in 15 immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 147

20 The translation product of this gene shares sequence homology with HNP36 protein, an equilibrative nucleoside transporter, which is thought to be important in gene transcription as well as serving as an important component of the nucleoside transport apparatus (See Accession No. 1845345). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

25 MSGQGLAGFFASVAMICAIASGSELSSESAGYFITACAVIILTIICYLGLPRLEFYR
YYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGVSVNSQPTNESHSIKAILK
NISVLAFSVCIFTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTFNIFDWLG
RSLTAVFMWPGKDSRWLPSWXLARLVFVPLLLLNIKPRRYLTVVFEHDAWFI
FFMAAFAFSNGYLASLCMCFGPKVKPAEAETAEPSPSSCVWWWHWGLFS

30 PSCSGQLCDKGWTEGLPASLPVCLLPLPSARGDPEWSGGFFF (SEQ ID
NO:705); MSGQGLAGFFASVAMICAIASGSELSSESAGYFITACAVIILTIIC
YLGLPRLEFYRYYQQLKLE GPGEQETKLDLISKGEEPRAGKEESGVSVNSQ
PTNESHSI (SEQ ID NO:706); SGVSVSNSQPTNESHSIKAILKNISVLAFSVCFI
FTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTFNIFDWLGRS (SEQ ID
35 NO:707), TIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLTFNIFDWLGRSLTAVF
MWPGKDSRWLPSWXLARLVFVPLLLLNIK PRRYLTVVFEHDA (SEQ ID
NO:708); FGPKVKPAEAETAEPSPSSCVWWWHWGLFSPSCSGQLCDK

GWTEGLPASLPVCLLPLPSARGDPEWSGGFFF (SEQ ID NO:709). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in eosinophils and aortic endothelium and to a lesser extent in umbilical vein endothelial cell and thymus.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential 10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having 15 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to HNP36 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of blood neoplasias and other hematopoietic disease.

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This gene is expressed primarily in breast cancer cell lines, thymus stromal cells, and ovary.

- 25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and female reproductive system diseases including breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell 30 type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard 35 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endocrine disorders. In addition, the tissue distribution in tumors of thymus, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are 5 useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 149

Translation product of this gene has homology to pmt1 and pmt 2, two conserved schizosaccharomyces pombe genes. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

DDDGFEIVPIEDPAKHRILDPEGGLALGAVIASSKKAKRDLIDNSFNRYTNEDEG

15 ELPEWFVQEEKQHRIRQLPVGKKEVEHYRKRWREINARPIXXXXXXXXXXXXX
XXXXXXXXLEQTRKKAEEAVVNTVDIXRTRES (SEQ ID NO:710);

DDDGFEIVPIEDPAKHRILDPEGGLALGAVIASSKKAKRDLIDNSFNRYT (SEQ

ID NO:711); KRWREINARPIXXXXXXXXXXXXXXXLEQTRKKAEE

AVVNTVDIXRTRES (SEQ ID NO:712). An additional embodiment is the

20 polynucleotide fragments encoding these polypeptide fragments (See Accession No. e1216734).

This gene is expressed primarily in retina and ovary and to a lesser extent in breast cancer cell, epididymus and osteosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal growth disorders, cancer and reproductive system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell 30 type(s). For a number of disorders of the above tissues or cells, particularly of the neural and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to 35 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 382 as residues: Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of reproductive system disease and cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 150

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MKDGRARTALTSSQPAHLCPENPLLHLKAAVKEKKRNKKKTIGSPKRIQS
PLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLS

10 SLQSDPAGCVRPPAPNLAGAVEFNDVKTLLREWITTISDPMEDILQVVKYCTD
LIEEKDLEKLDLVIKYMKRLMQQSVESVWNMAFDIFLDNVQVVLQQTYGSTLK
VT (SEQ ID NO:713); MKDKGRARTALTSSQPAHLCPENPLLHLKAAVKE
KKRNKKKTIGSPKRIQ (SEQ ID NO:714); KRIQSPLNNKLLNSPAKT
LPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPP
15 APNLAGAVEFNDVKTLLREWITTISDPM (SEQ ID NO:715);
TISDPMEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVE
SVWNMAFDIFLDNVQVVLQQTYGSTLKVT (SEQ ID NO:716). An additional
embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in 12 week embryo and to a lesser extent in
20 hemangiopericytoma and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth disorders and hemangiopericytoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circular and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 383 as residues: Leu-4 to Lys-11.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of growth disorders, hemangiopericytoma and other soft tissue tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 151

The translation product of this gene has been found to have homology to a human DNA mismatch repair protein PMS3. Preferred polypeptide fragments comprise 5 the following amino acid sequence: FCHDCKFPEASPAMNCEP (SEQ ID NO:717). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R95250).

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphoma, immunodeficiency diseases, and cancers resulting from genetic instability. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the 15 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to 20 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 384 as residues: Met-1 to Lys-6.

The tissue distribution in neutrophils and the sequence homology indicates that 25 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including 30 arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Furthermore, its homology to a known DNA repair protein would suggest gene may be useful in establishing cancer predisposition and prevention in gene therapy applications.

FEATURES OF PROTEIN ENCODED BY GENE NO: 152

35 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

- biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious diseases and lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders 5 of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 10 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of inflammation and infectious diseases.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 153

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPPVLRKKC
NFFCWDSSAHSLPLHPLSASCASAPACHASDTHLLYPSTRALCPSIFAWLVAPHS
20 VFRTNAPGPTPSSQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:720);
MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPPVLRKKCNFFCWDSSAH
SLPLHPLSASCASAPACHA (SEQ ID NO:721);FAWLVAPHSVFRTNAPGPTPS
SQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:722). An additional embodiment
is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infectious disease. Similarly, polypeptides and 30 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having 35 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 386 as residues: Ser-11 to Pro-17.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases and
5 inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed in multiple tissues including ovary, uterus, adipose tissue, brain, and the liver.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, uterine, ovarian, brain, and liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
20 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in
25 the treatment of the female reproductive system, obesity, and liver disorders, particularly cancer in the above tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 155

This gene maps to chromosome 3, and therefore, may be used as a marker in
30 linkage analysis for chromosome 3 (See Accession No. D87452).

This gene is expressed in multiple tissues including brain, aortic endothelial cells, smooth muscle, pituitary, testis, melanocytes, spleen, neutrophils, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders including immunodeficiencies, cancers of the brain and the female reproductive system, as well as cardiovascular disorders, such as

atherosclerosis and stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution suggest that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nervous system, including schizophrenia, neurodegeneration, neoplasia, brain cancer as well as cardiovascular and female reproductive disorders including cancer within the above tissues.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with the human gene encoding cytochrome b561 (See Accession No. P10897). Cytochrome b561 is a transmembrane electron transport protein that is specific to a subset of secretory vesicles containing catecholamines and amidated peptides. This protein is thought to supply reducing equivalents to the intravesicular enzymes dopamine-beta-hydroxylase and alpha-peptide amidase. Preferred polypeptides of the invention comprise the amino acid sequence:

MAMEGYWRFLALLGSALLVGFLSVIFALVVWLHYREGLGWDGSALEFNWHP

25 VLMVTGFVFIQGIAIVYRLPWTWKCSKLLMKSIIAGLNAVAAILAIISVVAVFE
NHNVNNIANMYSLHSWVGLIAVICYLLQLLSGFSVFLPWAPSLRAFLMPIHV
YSGIVIFGTIVIATALMGLTEKLIFSLRDPAYSTFPPEGVFVNTLGLLILVFGALIF
WIVTRPQWKRKPKEPNSTILHPNGGTEQGARGSMPAYSGNNMDKSDSEL
NSEVAARKRNLALDEAGQRSTM (SEQ ID NO:724); as well as antigenic fragments
30 of at least 20 amino acids of this gene and/or biologically active fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system and metabolism related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product or RNA of this gene is useful for treatment or diagnosis of immune system and metabolic diseases or conditions including Tay-Sachs disease, phenylketonuria, galactosemia, various porphyrias, and Hurler's syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 157

The translation product of this gene shares sequence homology with collagen which is important in mammalian development. This gene also shows sequence homology with bcl-2. (See Accession No. P80988.) Preferred polypeptide fragments comprise the amino acid sequence: PGRAGPSPGLSLQLPAEPGHPAGNLAPL TSRPQPLCRIPTAVPG (SEQ ID NO:725). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

This gene is expressed primarily in HL-60 tissue culture cells and to a lesser extent in liver, breast, and uterus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases, hereditary disorders involving the MHC class of immune molecules, as well as developmental disorders and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 390 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62.

The tissue distribution and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hereditary MHC disorders and particularly autoimmune disorders including rheumatoid arthritis, lupus, scleroderma, and dermatomyositis, as well as many reproductive disorders, including cancer of the uterus, and breast tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 158

10 This gene is expressed primarily in the amygdala region of the brain. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, particularly those effecting mood and personality. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and/or diagnosis of a variety of brain disorders, particularly bipolar disorder, unipolar depression, and dementia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 159

This gene is expressed in a variety of tissues and cell types including brain, smooth muscle, kidney, salivary gland and T-cells.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers of a variety of organs including brain, smooth muscle, kidney, salivary gland and T-cells and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

- of the above tissues or cells, particularly of the central nervous, urinary, salivary, digestive, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
- 5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, smooth muscle, and T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of

10 various neurological, and cardiovascular disorders, but not limited to cancer within the above tissues. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma,

15 immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with collagen which is thought to be important in cellular interactions, extracellular matrix formation,

20 and has been found to be an identifying determinant in autoimmune disorders. Moreover, this gene shows sequence homology with the yeast protein, Sls1p, an endoplasmic reticulum component, involved in the protein translocation process in Yeast *Yarrowia lipolytica*. (See Accession No. 1052828; see also J. Biol. Chem. 271, 11668-11675 (1996).) With mouse, this same region shows sequence homology with

25 the heavy chain of kinesin. (See Accession No. 2062607.) Recently, suppression of the heavy chain of kinesin was shown to inhibits insulin secretion from primary cultures of mouse beta-cells. (See Endocrinology 138 (5), 1979-1987 (1997).) Moreover, kinesin was found associated with drug resistance and cell immortalization. (See 468355.) Thus, it is likely that this gene also act as a genetic suppressor elements.

30 This gene is expressed primarily in the greater omentum and to a lesser extent in a variety of organs and cell types including gall bladder, stromal bone marrow cells, lymph node, liver, testes, pituitary, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the endocrine, gastrointestinal, and immunological systems, including autoimmune disorders and cancers in a variety of organs and cell types.

- Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or
- 5 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those
- 10 comprising a sequence shown in SEQ ID NO: 393 as residues: Asn-27 to Leu-47, Gln-81 to Lys-88, Asp-93 to Lys-102, Asn-107 to Leu-116, Met-129 to Glu-141, Glu-150 to Asp-157, Lys-176 to Glu-185, Glu-333 to Tyr-349, Cys-393 to Leu-403, Gln-423 to Gly-429.

The tissue distribution in within various endocrine and immunological tissues combined with the sequence homology to a conserved collagen motif indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various autoimmune disorders including, but not limited to, rheumatoid arthritis, lupus erythematosus, scleroderma, dermatomyositis. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 161

25 This gene has homology to the tissue inhibitor of metalloproteinase 2. Such inhibitors are vital to proper regulation of metalloproteins such as collagenases (See Accession No. P16368). In addition, this gene maps to chromosome 17, and therefore, may be used as a marker in linkage analysis for chromosome 17 (See Accession No. P16368).

30 This gene is expressed primarily in several types of cancer including osteoclastoma, chondrosarcoma, and rhabdomyosarcoma and to a lesser extent in several non-malignant tissues including synovium, amygdala, testes, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, various types of cancer, particularly cancers of bone and cartilage, as well as various autoimmune disorders. Similarly, polypeptides and antibodies directed

- to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- The tissue distribution in various cancers and the sequence homology to a 10 collagenase inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.
- 15

FEATURES OF PROTEIN ENCODED BY GENE NO: 162

This gene is homologous to the mitochondrial ATP6 gene and therefore is likely a homolog of this gene family (See Accession No. X76197).

- This gene is expressed primarily in brain tissue. 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, including Down's syndrome, depression, Schizophrenia, and epilepsy. Similarly, polypeptides and antibodies directed to these 25 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal 30 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain tissue indicates this gene is useful for diagnosis of various neurological disorders including, but not limited to, brain cancer.

- 35 Additionally the gene product may be used as a target in the immunotherapy of cancer in the brain as well as for the diagnosis of metabolic disorders such as obesity Tay-Sachs disease, phenylketonuria and Hurler's Syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in placenta, neutrophils, and microvascular endothelial cells and to a lesser extent in multiple tissues including brain, prostate,

5 spleen, thymus, and bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neutropenia and other diseases of the immune system. Similarly,

- 10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 15

The tissue distribution in placenta indicates that polynucleotides and

- 20 polypeptides corresponding to this gene are useful for diagnosis various female reproductive disorders. Additionally the gene product may be used as a target in the immunotherapy of various cancers. Because the gene is expressed in some cells of lymphoid and endocrine origin, the natural gene product may be involved in immune functions and metabolism regulation, respectively. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.
- 25

FEATURES OF PROTEIN ENCODED BY GENE NO: 164

This gene is expressed primarily in neutrophils, monocytes, bone marrow, and

30 fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders including, but not limited to, autoimmune

- 35 disorders such as lupus, and immunodeficiency disorders . Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

- of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
5 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various immune system tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various immunological disorders such as Hodgkin's lymphoma, arthritis,
10 asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 165

The translation product of this gene shares sequence homology with dystrophin which is thought to be defective in both Duchene and Becker Muscular Dystrophy.

- 15 Preferred polypeptide fragments comprise the following amino acid sequence:
MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNHLHSTETQTAGVIDRWELLQAQ
ALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELSTDITIELQ
IKKLKELQKAVDHRKAIILSINLCSPEFTQADSKESRDLQDRLXQMNGRWDRV
CSLLEEWRGLLQDALMQCQGFHEMSHGLLMLENIDRRKNEIVPIDSNLDAEIL
20 QDHHKQLMQIKHELLESQLRVASLQDMSCQLLVNAEGTDCLEAKEKVHVIGNR
LKLLLKEVSRHIKELEKLLDVSSSQQLSSWSSADELDTSGSVSPXSGRSTPNR
QKTPRGKCSLSQPGPSVSSPHSRSTKGGSDSSLSEPXPGRSGRGFLFRVLRAA
LPLQLLLLLIGLACLVPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
NO:726); MKLLGECSSSIDSVKRLEHKLKEEEESLPGFVNHLHSTETQTAGVIDR
25 WELLQAQALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELS
TDITIELQIK (SEQ ID NO:727); KLKELQKAVDHRKAIILSINLCSPEFTQADSK
ESRDLQDRLXQMNGRWDRVCSLLEWRGLLQDALMQCQGFHEMSHGLLML
ENIDRRKNEIVPIDSNLDAEILQDHHKQLMQIKHELLESQLRVASLQDMSCQL
(SEQ ID NO:728); QDMSCQLLVNAEGTDCLEAKEKVHVIGNRLKLLLKEVS
30 RHIKELEKLLDVSSSQQLSSWSSADELDTSGSVSPXSGRSTPNRQKTPRGKCS
LSQPGPSVSSPHS (SEQ ID NO:729); DSSLSEPXPGRSGRGFLFRVLRAAL
PLQLLLLLIGLACLVPMSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
NO:730). Also preferred are polynucleotide fragments encoding these polypeptide
fragments. Furthermore, this gene maps to chromosome 6, and therefore, may be used
35 as a marker in linkage analysis for chromosome 6 (See Accession No. N62896).

This gene is expressed in numerous tissues including the heart, kidney, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, musculoskeletal disorders including Muscular Dystrophy and

- 5 cardiovascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscle tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded 10 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution and homology to dystrophin indicates that 15 polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of Muscular Dystrophy and other muscle disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human cerebellum.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the central nervous system, including Alzheimer's Disease, Parkinson's Disease, ALS, and mental illnesses. Similarly, polypeptides and antibodies 25 directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual 30 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 399 as residues: Pro-20 to Gly-26, Leu-37 to Pro-42, His-57 to Gly-63.

- 35 The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the central nervous system and may protect or

enhance survival of neuronal cells by slowing progression of neurodegenerative diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 167

- 5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MKLLICGNYLAPSHSESSRRCCLLCFYPLCLEINFGMKVFLSMPFLVLFQ

SLIQED (SEQ ID NO:731). Polynucleotides encoding such polypeptides are also provided. This gene is believed to reside on chromosome 15. Therefore polynucleotides derived from this gene are useful in linkage analysis as chromosome 15 markers.

10 This gene is expressed primarily in human testes tumor and to a lesser extent in normal human testes.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the testes, particularly cancer, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of 20 the male reproductive tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily 25 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of testicular diseases including cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 168

- 30 This gene is expressed primarily in fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, conditions affecting hematopoietic development and metabolic diseases.

- 35 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

- hepatic system, and fetal hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,
- 5 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 401 as residues: His-7 to Trp-17, Leu-19 to Lys-27, Pro-33 to Gly-44, Lys-68 to Gly-74, Lys-85 to Cys-95.

The tissue distribution indicates that the protein products of this gene are useful
10 for treatment/diagnosis of diseases of the developing liver and hematopoietic system, and act as a growth differentiation factor for hematopoietic stem cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The polypeptide encoded by this gene is believed to be a membrane bound
15 receptor. The extracellular domain of which is expected to consist of the following amino acid sequence:

RILLVKYSANEENKYDYLPTTVNVCSELVKLVFCVLVSFCVIKKDHQSRLNKY
ASWKEFSDFMKWSIPAFLYFLDNLIVFYVLSYLPAMAVIFSNFSIITTALLFRIV
LKXRLNWIQWASLLTLFLSIVALTAGTKTLQHNLAGRGFHDAFFSPNSCLL
20 FRNECPKDNCTAKEWTFPEAKWNTTARVFSHIRLGGMGHVLIIVQCFISSMANI
YNEKILKEGNQLTEXIFIQNSKL^YFFGILFNGLTGLQRSNRDQIKNCGFFYGH
S (SEQ ID NO:732). Thus, preferred polypeptides encoded by this gene comprise the extracellular domain as shown above. It will be recognized, however, that deletions of either end of the extracellular domain up to the first cysteine from the N-terminus and
25 the first cysteine of the C-terminus, is expected to retain the biological functions of the full-length extracellular domain because the cysteines are thought to be responsible for providing secondary structure to the molecule. Thus, deletions of one or more amino acids from either end (or both ends) of the extracellular domain are contemplated. Of course, further deletions including the cysteines are also contemplated as useful as such
30 polypeptides is expected to have immunological properties such as the ability to evoke and immune response. Polynucleotides encoding all of the foregoing polypeptides are provided.

This gene is expressed primarily in human osteoclastoma and to a lesser extent in hippocampus and chondrosarcoma.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancers, particularly those of the bone and connective tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system,
5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of cancers of the bone and connective tissues, and may act as growth factors for cells involved in bone or connective tissue growth.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 170

Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

NSVPNLQTLAVLTEAIGPEPAIPRXPREPPVATSTPATPSAGPQPLPTGTV
20 LVPGGPAPPCLGEAWALLPPCRPSLTSCFWSPRPSPWKETGV (SEQ ID NO:733). Polynucleotides encoding such polypeptides are also provided herein.

This gene is expressed primarily in hematopoietic progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the blood including cancer and autoimmune disorders.
25 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the blood/circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 403 as residues: Gln-4 to His-10, Pro-25 to His-32.
35

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of diseases involving growth differentiation of hematopoietic cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 171

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequences: ALQLAFYPDAVEEWLEENVHPSLQRLQXLLQDLSEVSAPP (SEQ ID NO:734); and/or CHPPALAGTLLRTPEGRAHARGLLLEAGGA (SEQ ID NO:735). Polynucleotides encoding such polypeptides are also provided. The protein product of this gene shares sequence homology with metallothioneins. Thus, polypeptide encoded
10 by this gene are expected to have metallothioneine activity, such activities are known in the art and described elsewhere herein.

This gene is expressed primarily in kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the kidney including cancer and renal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system,
20 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
25 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 404 as residues: Ser-47 to Gln-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the kidney including kidney failure.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine,

5 synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 405 as residues: Gln-31 to Thr-43, Gly-51 to Ser-58, Pro-65 to Pro-72.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of developmental problems with fetal tissue. The gene may be involved in vital organ development in the early stage, especially hematopoiesis, cardiovascular system, and neural development.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 173

The translation product of this gene shares sequence homology with TGN38, an integral membrane protein previously shown to be predominantly localized to the trans-Golgi network (TGN) of cells.

20 This gene is expressed primarily in developing embryo and to a lesser extent in cancer tissues including lymphoma, endometrial, prostate and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and cancer. Similarly, polypeptides and 25 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual 30 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 406 as residues: His-65 to Ser-72, Pro-82 to Gly-91, Pro-98 to Glu-118, Ser-126 35 to Gly-166, Pro-180 to Asp-188, Tyr-209 to Lys-214, Gln-220 to Leu-228.

The tissue distribution and homology to an integral membrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for

diagnosis of cancers and developmental abnormalities where aberrant expression relates to an abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 174

5 The translation product of this gene shares sequence homology with a dnaJ heat shock protein from *E. coli* which is allelic to sec63, a gene that affects transit of nascent secretory proteins across the endoplasmic reticulum in yeast.

This gene is expressed primarily in Hodgkin's lymphoma and to a lesser extent in testes.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to
20 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 407 as residues: Thr-13 to Trp-21, Arg-74 to Asp-81.

The tissue distribution and homology to dnaJ indicates that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic for cancer including Hodgkin's lymphoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 175

This gene is expressed primarily in endothelial cells and to a lesser extent in
30 bone marrow stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving angiogenic abnormalities including diabetic
35 retinopathy, macular degeneration, and other diseases including arteriosclerosis and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

- type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or 5 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treating diseases where an increase or decrease in angiogenesis is indicated and as a 10 factor in the wound healing process.

FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with MAT8 (mouse) which is thought to be important in regulating chloride conductance in cells 15 (particularly in the breast) by modulating the response mediated by cAMP and protein kinase C to extracellular signals.

This gene is expressed primarily in amniotic cells and hematopoietic cells including macrophages, Neutrophils, T cells, TNF induced aortic endothelium and to a lesser extent in testes, TNF induced epithelial cells, and smooth muscle. 20

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory responses mediated by T cells, macrophages, and/or neutrophils particularly those involving TNF, and also cancer. Similarly, polypeptides 25 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 409 as residues: Thr-19 to Ala-33, Leu-54 to Asp-82, Pro-89 to Ala-97, Pro-100 35 to Lys-125, Ser-127 to Phe-135, Gly-164 to Leu-169, Cys-173 to Arg-178.

The tissue distribution and homology to mat-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for modifying inflammatory

responses to cytokines such as TNF and thus modifying the duration and/or severity of inflammation. Polynucleotides and polypeptides derived from this gene are thought to be useful in the diagnosis and treatment of cancer.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 177

- This gene is expressed primarily in endothelial cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 10 not limited to, vascular restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases associated with vascular response to injury such as vascular restenosis following angioplasty..

FEATURES OF PROTEIN ENCODED BY GENE NO: 178

- One embodiment of the claimed invention comprises:
- 25 MRPDWKAGAGPGGPPQKPAPSSQRKPPARPSAAAAIAVAAAEEERRLRQRN
RLRLEEDKPAVERCLEELVFGDVENDEDALLRRRGPRVQEHEDESGDSEVENEA
KGNFPQKKPVVVDEEDEDDEEMVDMMNNRFRKDMMKNASESKLSKDNLKK
RLKEEFQHAMGGVPAWAETTKRKTSSDDESEEEDEDLLQRTGNFISTSTSLPRG
ILKMKNCQHANAERPTVARISICAVPSRCTDCDGCWD (SEQ ID NO:737); or
- 30 CLEELVFGDVENDEDALLRRRGPRVQEHEDESGDSEVENEA
KGNFPQKKPV
WVDEEDEDDEEMVDMMNNRFRKDMMKNASESKLSKDNLKKRLKEEFQHAMG
GVPAWAETTKRKTSSDDESEEEDEDLLQRTGNFISTSTSLPRGILKMKNCQHA
NAERPTVARISICAVPSRCTDCDGC (SEQ ID NO: 738). LKEKIVRSFEVSPDG
FLLINGIAGYLHLLAMKTKELIGSMKINGRVAASTFSSDSKKVYASSGDGEVYV
- 35 WDVSNSRKCLNRVDEGSLYGLSIATSRNGQYYACGSNCGVVNIYNQDSCLQE
TNPKPIKAIMNLVTGVTSLTFNPTEILAIASEKMKEA
VRLVHLPSCTVFSNFPVI
KNKNISHVHTMDFSPRSGYFALGNEKGKALMYRLHHYSDF (SEQ ID NO:739);

and/or KINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSL
YGLSIATSRNGQYVACGSNCGVNIYNQDSCLQETNPKPIKAIMNLVTGVTSLT
FNPTTEILAIASEKMKEAVRLVHLPSCTVFSNFPVIKNKNISHVHTMDFSPRSG
YFALGNEKGKAL (SEQ ID NO:740).

5 This gene is expressed primarily in epididymus and endometrial tumors and to a lesser extent in T cell lymphoma and cell lines derived from colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 10 not limited to, tumors of the reproductive organs including testis and endometrial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may 15 be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a 20 sequence shown in SEQ ID NO: 411 as residues: Ser-67 to Lys-72, Val-87 to Leu-93, Tyr-128 to Pro-141, Asp-204 to Gly-210.

The tissue distribution indicates that the protein products of this gene are useful for treating tumors of the endometrium or epithelial tumors of the reproductive system.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 179

Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MRILQLILLALATGLVGGETRIIKGFECKLHSQPWQAALFEKTRLLCGATLIAPR
WLLTAAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDH
30 RNDIMLVKMASPVSIWAVRPLTLSSRCVTAGTSCSFPAAGAARPDPSYACLTPC
DAPTSPLSTRSVRTPTPATSQTPWCVPACRGARTPARVTPGALWSVTSLFKA
LSPGARIRVRSPESLVSTRKSANMWTSRRR (SEQ ID NO:741); ETRIIKGFEC
KLHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNLQKEE
GCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSIWAVRPLTLSSR
35 CVTAGTSCSFPAAGAARPDPSYACLTPCDAPTSPLSTRSVRTPTPATSQTPWCVP
ACRGARTPARVTPGALWSVTSLFKALSPGARIRVRSPESLVSTRKSANMWTS

SRRR (SEQ ID NO:742); or CKLHSQPWQAALFEKTRLLCGATLIAPRWLLT
AAHCLKPRYIVHLGQHNLQKEEGCEQRTATESFPHPGFNS
(SEQ ID NO:743). The translation product of this gene shares sequence homology
with neuropsin a novel serine protease which is thought to be important in modulating
5 extracellular signaling pathways in the brain. Owing to the structural similarity to other
serine proteases the protein products of this gene are expected to have serine protease
activity which may be assayed by methods known in the art and described elsewhere
herein.

This gene is expressed primarily in endometrial tumor and to a lesser extent in
10 colon cancer, benign hypertrophic prostate, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cancers of the endometrium or colon and benign hypertrophy of the
15 prostate. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the tissue(s)
or cell type(s). For a number of disorders of the above tissues or cells, particularly of
the urogenital or reproductive systems, expression of this gene at significantly higher or
lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
comprising a sequence shown in SEQ ID NO: 412 as residues: Gly-12 to Ser-22, Pro-
25 34 to Ser-53.

The tissue distribution and homology to serine proteases indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diagnosing
or treating hyperproliferative disorders such as cancer of the endometrium or colon and
hyperplasia of the prostate.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 180

Preferred polypeptide encoded by this gene comprise the following amino acid
sequence: VLQGRYFSPILEMRRLRPEGXXNLPGGSRAQKEPRQDLTLVLWPHC
PHFAMTRS YVPTKQCMVQGSFYCIFIFKGPVQNWC (SEQ ID NO:744).

35 Polynucleotides encoding such polypeptide are also provided.

This gene is expressed primarily in fetal brain

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, identifying and expanding stem cells in the CNS. Similarly, polypeptides 5 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 10 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful 15 for detecting and expanding stem cell populations in the (or of the) central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in early stage human brain and a stromal cell 20 line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities of the CNS. Similarly, polypeptides and 25 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, 30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 414 as residues: Gln-42 to Gln-47, Gln-54 to Pro-60.

35 The tissue distribution indicates that the protein products of this gene play a role in the development of the central nervous system. Therefore this gene and its products

are useful for diagnosing or treating developmental abnormalities of the central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 182

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MPIIDQVNPELHDFMQSAEVGTIFALSWLITWFGHVLSDFRHVVRLYDF
FLACHPLMPIYFAAVIVLYREQEVLDCDCDMASVHHLLSQIPQDLPYETLISRXE
TFLFSFPHPNLLGRPLPNSKLRGRQPLLSKTL SWHQPSRGLIWCCGSGXRGLL

10 RPEDRTKDVLT KPTNRFVKLAVMGLTVALGAAALAVVKA LEWAPKFQLQL
FP (SEQ ID NO:745); or CPEFFIPATLPCPFVFAFTSEASSRAYLTQRGPGLAQ
NLMPLPVGFWMGS LPPPWCWRKWVSEACSCFC (SEQ ID NO:746) These
polypeptides are structurally similar to various TGF-beta family members. Thus, this
polypeptide is expected to have a variety of activities in the modulation of cell growth
15 and proliferation.

This gene is expressed primarily in osteoclastoma, microvascular endothelium,
and bone marrow derived cell lines.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, hematological diseases particularly involving aberrant proliferation of
stem cells. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the tissue(s)
or cell type(s). For a number of disorders of the above tissues or cells, particularly of
25 the immune system, expression of this gene at significantly higher or lower levels may
be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a
sequence shown in SEQ ID NO: 415 as residues: Ser-33 to Ala-39.

The tissue distribution indicates that the protein products of this gene is useful
for treating disorders of the progenitors of the immune system. Applications include in
vivo expansion of progenitor cells, ex vivo expansion of progenitor cells, or the
35 treatment of tumors of the circulatory system, such as lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 183

This gene maps to chromosome 17 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 5 GFGSVSAAGRRSGGTWQPVQ (SEQ ID NO:747); PGGLAVGSRWWSRSLT
(SEQ ID NO:748); LEPSRQRRPQQRGTSRPETDQRAKCWRQL (SEQ ID
NO:749); and/or VCLRCQNRMEN (SEQ ID NO:750). In further specific
embodiments, polypeptides of the invention comprise the sequence: MAACTARRPGR
GQPLVVVPVADXGPVAKAALCAAXAGAFSPASTTTRRLSSRNRPPEGKVLETV
10 GVFEVPKQNGKYETGQLFLHSIFGYRGVVLFPWQARLXDRDVASAAPEKAEN
PAGHGSKEVKGKTHYYQVLIDARDCPHISQRSQTEAVTFLANHDDSRALYAIP
GLDYVSHEDILPYTSTDQVPIQHELPFLLYDQTKAPPVARETLRAWQEKNH
PWLELSDVHRETENIRVTVIPFYMGMREAQNSHVYWWRYCIRLENLDSDVVQ
LRERHWRIFSLSGTLETVRGRGVVGREPVLSEQPAFQYSSHVSLQASSGHMW
15 GTFRFERPDGSHFDVRIPPFSLSNKDEKTPPSGLHW (SEQ ID NO:751);
MAACTARRPGRGQPLVVVPVADXGPVAKAALCAA (SEQ ID NO:752);
VLETVGVFEPKQNGKYETGQLFLHSIFGYRGVVL (SEQ ID NO:757);
GLDYVSHEDILPYTST (SEQ ID NO:758); DVHRETENIRVTVIPFYM (SEQ ID
NO:759); WWRYCIRLENLDSDVVQLRER (SEQ ID NO:760); and/or PAFQYSS
20 HVSLQASSGHMWGTFRFER (SEQ ID NO:761). Polynucleotides encoding these
polypeptides are also encompassed by the invention.

This gene is expressed primarily in gall bladder, prostate, and fetal brain, and to a lesser extent in a few tumor and fetal tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, growth related disorders such as cancers. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
30 of the above tissues or cells, particularly of the prostate, gall bladder, and fetal brain,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
35 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth-related disorders, such cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 184

In specific embodiments, polypeptides of the invention comprise the sequence: SLCCPEGAEGC (SEQ ID NO: 762) and/or QLKKTHYDRPCP (SEQ ID NO: 763). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in stromal cell, tonsil, and glioblastoma and to a lesser extent in some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

15 not limited to, immune and inflammatory disorders and glioblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, tonsil, and glioblastoma expression of this gene at significantly higher or lower levels

20 may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, it is believed that the

25 product of this gene regulates pancreatic cell differentiation into beta cells. Accordingly, polynucleotides and polypeptides of the invention are useful in the treatment of insulin-dependent diabetes mellitus and associated conditions e.g. pancreatic hypofunction and the prevention, as well as the treatment of undifferentiated type pancreatic cancers.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 417 as
30 residues: Pro-27 to Ala-32.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune and inflammatory disorders and glioblastoma.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in hepatocellular carcinoma and to a lesser extent in other tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver diseases. Similarly, polypeptides and antibodies directed to these 5 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another 10 tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 418 as residues: Gly-32 to Lys-39.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in hippocampus and to a lesser extent in other tissues.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential 25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, 30 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

35 **FEATURES OF PROTEIN ENCODED BY GENE NO: 187**

This gene is expressed primarily in bone cancer and hippocampus and to a lesser extent in osteoclastoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone-related disorders and neuronal diseases. Similarly, polypeptides 5 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, osteoclast, and hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily 10 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides 15 corresponding to this gene are useful for diagnosis and treatment of bone-related disorders and neuronal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 188

This gene maps to chromosome 4 and therefore polynucleotides of the invention 20 can be used in linkage analysis as a marker for chromosome 4.

This gene is expressed primarily in neuronal tissues such as hippocampus, spinal cord, and hypothalamus and to a lesser extent in a few other tissues such as ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above 30 tissues or cells, particularly of the neuronal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 35 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 189

This gene maps to chromosome 10, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 10.

- 5 This gene is expressed primarily in neuronal tissues and immune tissues, and to a lesser extent in a few other tissues such as skin tumor, lung etc.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
- 10 not limited to, neuronal and immune-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal and immune-related tissues, expression of this gene at significantly higher or lower levels may be routinely detected
- 15 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
- 20 NO: 422 as residues: Pro-19 to Asp-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal and immune-related disorders.

25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 190**

The translation product of this gene shares sequence homology with human N33, a gene located in a homozygously deleted region of human metastatic prostate cancer which is thought to be important in prevention of prostate cancer. In specific embodiments, polypeptides of the invention comprise the sequence:

- 30 AQRKKEMVLSEKVSQLM**EWTNKRPVIRMNGDKFRRLVKAPRNYSVIVMFTA**
LQLHRQCVVCKQADEFQILANSWRYSSAFTNRIFAMVDFDEGSDVFQMLNM
NSAPTFINFPAKGKPKRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPNMA
ARWRFWCVSVT (SEQ ID NO:765); MVVALLIVCDVPSAS (SEQ ID NO:766);
AQRKKEMVLSEKVSQ**L (SEQ ID NO:767); MEWTNKRPVIRMNGDKF (SEQ**
- 35 ID:768); RRLVKAPPNYSVIVMFTALQLHRQCVVCKQADEFQILANSWRY
SSAFTNRIFFA (SEQ ID NO:769); MVDFDEGSDVFQMLNMNSAPTFINFPAK
GKP (SEQ ID NO:770); KRGDTYELQVRGFSAEQIARWIADRTDVNIRVIRPPN

(SEQ ID NO:771); and/or YAGPLMLGLLLAVIGGLVYLRRVIWNFSLIKLDGLLQL CVLCLL (SEQ ID NO:772). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in infant adrenal gland prostate cell line and to
5 a lesser extent in a few other tissues like liver, smooth muscle etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate cancer and endocrine disorders. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and adrenal gland, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 423 as residues: Pro-34 to Gly-43, Arg-113 to Pro-120.

20 The tissue distribution and homology to N33 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for prostate cancer and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 191

25 This gene is expressed primarily in T cell and to a lesser extent in fetal lung. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
30 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid)
35 or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 424 as residues: Trp-3 to Phe-9.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 192

This gene maps to chromosome 6, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 6. Neural activity and neurotrophins induce synaptic remodeling in part by altering gene expression. This

- 10 gene is believed to be a glycosylphosphatidylinositol-anchored protein encoded by a hippocampal gene and to possess neural activity. This molecule is believed to be expressed in postmitotic-differentiating neurons of the developing nervous system and neuronal structures associated with plasticity in the adult. Message of this gene is believed to be induced by neuronal activity and by the activity-regulated neurotrophins
15 BDNF and NT-3. The product of this gene is believed to stimulate neurite outgrowth and arborization in primary embryonic hippocampal and cortical cultures and to act as a downstream effector of activity-induced neurite outgrowth. In specific embodiments, polypeptides of the invention comprise the sequence: DAVFKGFSDCLLKLGDS (SEQ ID NO:773); CQEGAKDMWDKLRKESKNLN (SEQ ID NO:774);
20 VLLVSLSAALATWLSF (SEQ ID NO:775); MGLKLNGRYISLILAVQIAYLVQAVR AAGKCDAVFKGFSDCLLKLGDS (SEQ ID NO:776); PAAWDDKTNIKTVCTYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLNQGSLFELCGSGNGAAGSL LPAFPVLLVSLSAALATWLSF (SEQ ID NO:777); and/or MGLKLNGRYISLILA VQIAYLVQAVRAAGKCDAVFKGFSDCLLKLGDSXXXXXPAAWDDKTNIKTVC
25 TYWEDFHSCVTALTDCQEGAKDMWDKLRKESKNLNQGSLFELCGSGNGAA GSLLPAFPVLLVSLSAALATWLSF (SEQ ID NO:778). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in human placenta, endometrial tumor and tissues of the central nervous system (CNS).

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, relating to reproductive disorders, cancers and neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
35 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and neurological disorders, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 425 as residues: Asp-47 to Asp-63, His-75 to Tyr-80, Pro-83 to Tyr-89.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive disorders such as endometrial tumors. Expression of this gene in tissues of the CNS and its strong homology to Neuritin suggest that the protein product from this gene may also be used in the treatment and diagnosis of neurological disorders and in the regeneration of neural tissues, e.g., following injury.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 193

The translation product of this gene shares sequence homology with tenascin which is thought to be important in development. The translation product of this gene is believed to be a ligand of the fibroblast growth factor family. FGF ligand activity is known in the art and can be assayed by methods known in the art and disclosed elsewhere herein.

This gene is expressed primarily in endometrial tumors, and other types of tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 426 as residues: Gly-29 to Glu-34, Arg-71 to Arg-76, Thr-176 to Cys-182, Gly-184 to Glu-199.

The tissue distribution and homology to tenascin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 194

In specific embodiments, polypeptides of the invention comprise the sequence: MNSAAGFSHLDRRERVLKLGESFEKQPRCASTLC (SEQ ID NO:779).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal human lung and neutrophils.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lung development and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
20 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution in fetal lung and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
25 and treatment of lung and immunity related diseases, for example, lung cancer, viral, fungal or bacterial infections (e.g. lesions caused by tuberculosis), inflammation (e.g. pneumonia), metabolic lesions etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 195

- 30 This gene is expressed primarily in breast lymph node.
Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunal disorders. Similarly, polypeptides and antibodies directed to
35 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

5 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunal disorders.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 196

This gene maps to chromosome 5 and accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 5. The translation product of this gene shares sequence homology with human M-phase phosphoprotein 4 which is thought to be important in phosphorylation and signal transduction processes. In

15 specific embodiments, polypeptides of the invention comprise the sequence:

TIYPTEELQAVQKIVSITERALKLVSD (SEQ ID NO:780); RALKGVLRV
GVLAKGLLRGDRNVNLVLLC (SEQ ID NO:781); ALAALRHAKWFQARAN
GLQSCVIIIRILRDLCQRVPTWS (SEQ ID NO:782); GDALRRVFECIISSGIIL (SEQ
ID NO:783); LAFRQIHKVLGMDPLP (SEQ ID NO:784); and/or TIYPTEELQAVQ
20 KIVSITERALKVSDLSEHEKNKNKEGDDKKEGGKDRALKGVLRGVVLAKG
LLLRGDRNVNLVLLCSEKPSKTLSSRIAENLPKQLAVISPEKYDIKCAVSEAAIIL
NSCVEPKMQVTITLTSPIIREENMREGDVTSGVKDPPDVLDRKCLDALAALR
HAKWFQARANGLQSCVIIIRILRDLCQRVPTWSDFPSWAMELLVEKAISSASSP
QSPGDALRRVFECIISSGIILKGSPGLLDPCEKDPFDTLATMTDQQREDITSSAQFA
25 LRLLAFRQIHKVLGMDPLPQMSQRFNHHNNRKKRSDGVDGFQEAEGKKDKK
DYDNF (SEQ ID NO:785). Polynucleotides encoding these polypeptides are also
encompassed by the invention.

This gene is expressed primarily in Human Hippocampus and to a lesser extent in Prostate, Human Frontal Cortex.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders related to reproductive system and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and nervous system, expression of this gene at significantly higher or lower

levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to human M-phase phosphoprotein 4 indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and nervous system disorders.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 197

In specific embodiments, polypeptides of the invention comprise the sequence:

MGSQHSAAARPSSCRRKQEDDRDG (SEQ ID NO:786);

LLAEREQEEAIAQFPYVEFTGRDSITCLTC (SEQ ID NO:787); and/or

QGTGYIPTEQVNELVALIPHSDQRLRPQRTKQYV (SEQ ID NO:788).

15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Primary Breast Cancer and to a lesser extent in Human Adult Spleen, Hodgkin's Lymphoma I, Salivary Gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and immunal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and immune system, expression of 25 this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the 30 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 430 as residues: Ser-126 to Gly-138.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and immunal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 198

This gene is expressed primarily in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, blood cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of blood cell disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 199

This gene is expressed primarily in Human Ovary and Synovia and to a lesser extent in Human 8 Week Whole Embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 200

This gene maps to chromosome 8 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 8. The translation product of this gene shares limited sequence homology with collagen proline rich domain.

5 This gene is expressed primarily in CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are 10 not limited to, neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., 15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 433 as residues:

20 Pro-35 to Asp-41.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological diseases.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 201

Translation product of this gene shares homology with a mammalian histone H1a protein. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: ARLNVGRESLKREMLKSQGVKVSESPMGAR HSSWPEGAAFCKKVQGAQMFPPIR (SEQ ID NO:789); ARLNVGRESLKR 30 EML (SEQ ID NO:790); LKSQGVKVSESPMGARHSSW (SEQ ID NO:791); AFCKKVQGAQMFPPIR (SEQ ID NO:792). An additional embodiment is the polynucleotide fragments encoding these polypeptide (See Accession No. pirlS24178) fragments.

This gene is expressed primarily in neutrophils.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in vital immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such 15 as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 202

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as 20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above 25 tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for 35 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 203

This gene is expressed primarily in Neutrophils.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious disorders, immune disorders, and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For 10 a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 15 the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 436 as residues: Thr-31 to Lys-36.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of infectious 20 disorders, immune disorders, and cancers. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including 25 arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 204

- This gene maps to chromosome 16 and therefore polynucleotides of the invention can be used in linkage analysis as markers for chromosome 16. The 30 translation product of this gene shares sequence homology with lactate dehydrogenase which is thought to be important in lactate metabolism.

This gene is expressed primarily in human tonsils and to a lesser extent in Spleen, and Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as 35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, infectious disorders, and cancers. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune

disorders, infectious disorders, and cancers, expression of this gene at significantly

- 5 higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include
- 10 those comprising a sequence shown in SEQ ID NO: 437 as residues: Gly-7 to Ser-12.

The tissue distribution and homology to lactate dehydrogenase gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, infectious disorders, and cancers.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 205

The translation product of this gene shares sequence homology with Gcap1 protein which is developmentally regulated in brain.

This gene is expressed primarily in placenta and endometrial tumor and to a lesser extent in several other tumors.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vasculogenesis/angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological
- 25 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
- 30 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Gcap1 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorder or dysfunction of vascular system of tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 206

In specific embodiments, polypeptides of the invention comprise the sequence MPYAQWLAENDRFEEAQKAFHKAGRQREA (SEQ ID NO:799);

VQVLEQLTNNAVAESRFNDAAYYYWMLSMQCLDIAQD (SEQ ID NO:794);

- 5 PAQKDTMLGKFYHFQRLAELYHGYHAIHRHTEDP (SEQ ID NO: 795);
FSVHRPETLFNISRFLLHSLPKDTPSGISKVKILFT (SEQ ID NO:800);
LAKQSKALGAYRLARHAYDKLRGLYIP (SEQ ID NO:796); ARFQKSIELG
TLTIRAKPFHDSEELVPLCYRCSTNN (SEQ ID NO: 797); and/or PLLNNLGNVC
INCRQPFIFSASSYDVLHLVEFYLEEGITDEEAISLIDLEVLRPKRDDRQLEICKQQ
- 10 LPDSCG (SEQ ID NO:798). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of male reproductive and endocrine disorders.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 207

This gene is expressed in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lung diseases such as cystic fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

- of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
- 5 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 440 as residues: Tyr-49 to Cys-54.
- The tissue distribution indicates that polynucleotides and polypeptides
- 10 corresponding to this gene are useful for detection and treatment of disorders associated with developing lungs particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of lung tumors since the gene may be involved in the regulation of cell division,
- 15 particularly since it is expressed in fetal tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	First SEQ ID NO: Y	Last AA Sig Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	First AA of ORF
1	HLHDS67	97979 03/27/97	Uni-ZAP XR	11 2526	427 2526	458 458	234 1	30	31	30	31	30	30	115
2	HLRHDZ58	97979 03/27/97	Uni-ZAP XR	12 1131	1 1131	129 129	235 1	14	15	14	15	14	15	102
3	HLMMMJ13	97979 03/27/97	Lambda ZAP II	13 941	39 941	62 62	236 1	44	45	44	45	35	36	41
3	HLMMMJ13	97979 03/27/97	Lambda ZAP II	218 941	39 941	245 245	441 1	19	20	19	20	20	20	42
4	HLTE25	97979 03/27/97	Uni-ZAP XR	14 843	1 843	155 155	237 1	18	19	18	19	18	19	36
5	HMSJX24	97979 03/27/97	Uni-ZAP XR	15 1018	1 1018	90 90	238 1	28	29	28	29	28	29	127
6	HNFED65	97979 03/27/97	Uni-ZAP XR	16 661	1 661	76 76	239 1	23	24	23	24	23	24	66
7	HNHDX07	97979 03/27/97	Uni-ZAP XR	17 553	1 553	106 106	240 1	21	22	21	22	21	22	68
8	HNHGCC82	97979 03/27/97	Uni-ZAP XR	18 869	1 869	101 101	241 1	21	22	21	22	21	22	44
9	HNHGO09	97979 03/27/97	Uni-ZAP XR	19 959	1 959	176 176	242 1	21	22	21	22	21	22	44
10	HOUBE18	97979 03/27/97	Uni-ZAP XR	20 1446	1 1446	101 101	243 1	27	28	27	28	27	28	50
11	HOUDL69	97979 03/27/97	Uni-ZAP XR	21 1471	579 1460	692 692	244 1	31	32	31	32	31	32	42

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA of Signal Pep	5' NT of AA of SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HPMFI71	97979 03/27/97	Uni-ZAP XR	22	1402	242	1402	401	245	1	32	33	60
13	HPMGQ55	97979 03/27/97	Uni-ZAP XR	23	1047	1	1047	164	164	246	1	26	27
14	HPQAC69	97979 03/27/97	Lambda ZAP II	24	990	1	988	82	82	247	1	20	21
15	HPTBB03	97979 03/27/97	Uni-ZAP XR	25	1208	350	1173	398	398	248	1	29	30
16	HPTWA66	97979 03/27/97	pBluescript	26	1922	1381	1922	24	24	249	1	33	34
16	HPTWA66	97979 03/27/97	pBluescript	219	575	1	575	148	148	442	1	22	23
17	HPTWC08	97979 03/27/97	pBluescript	27	1951	1422	1874	219	219	250	1	19	20
18	HRGCZ46	97979 03/27/97	Uni-ZAP XR	28	3989	2635	3989	2748	251	1	16	17	39
19	HSAVU34	97979 03/27/97	Uni-ZAP XR	29	3735	2966	3735	272	272	252	1	30	31
19	HSAVU34	97979 03/27/97	Uni-ZAP XR	220	3018	1929	3018	26	26	443	1	1	2
20	HSDFW61	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	30	1667	59	1625	138	138	253	1	32	33
21	HSDGP60	97974 04/04/97	Uni-ZAP XR	31	1408	1	1408	285	285	254	1		20

163

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
22	HSOAJ55	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	32	2031	1273	2031	1285	1285	255	1	29	30
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	33	971	13	971	91	91	256	1	19	20
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	221	968	8	968	86	86	444	1	20	21
24	HSXAM05	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	34	1792	369	1792	470	470	257	1	26	27
25	HSXAS67	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	35	896	1	896	96	96	258	1	32	33
26	HTDAF28	97974 04/04/97 209080 05/29/97	pSport1	36	912	1	912	38	38	259	1	22	23
27	HTEGQ64	97974	Uni-ZAP XR	37	1382	67	1382	271	271	260	1		25

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
28	HTGEU09	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	38	872	1	872	74	74	261	1	18	19
29	HTOAM21	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	39	812	1	812	41	41	262	1	30	31
30	HTPBW79	209511 12/03/97	Uni-ZAP XR	40	1515	118	1507	302	302	263	1	24	25
30	HTSEV09	97974 04/04/97 209080 05/29/97	pBluescript	222	1404	1	1265	92	92	445	1	19	20
31	HJPCD40	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	41	704	22	704	117	264	1	18	19	127
32	HTWBY48	97974 04/04/97 209080 05/29/97	pSport1	42	1094	1	1094	32	32	265	1	34	35
33	HTWCI46	97974 04/04/97	pSport1	43	1821	892	1647	56	56	266	1	26	27

165

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	S' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HCDAR68	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	50	1487	181	1455	325	325	273	1	35	36
41	HCE3J79	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	51	1328	251	1328	525	525	274	1		21
42	HMDAN54	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	52	1856	725	1853	928	928	275	1	33	34
43	HCECA49	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	53	1558	310	1408	393	393	276	1		1
44	HCEEC15	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	54	948	1	948	9	9	277	1	23	24
45	HCESF40	97974 04/04/97 209080 05/29/97	pBluescript	55	990	99	990	193	193	278	1	32	33

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID NO: Y	First AA of Signal Pep	5' NT of AA SEQ ID NO: Z	First AA of Signal Pep	Last AA of Signal Pep	First AA of Secreted Portion	Last AA of Secreted Portion	First AA of ORF
45	HCESF40	97974 04/04/97 05/29/97	pBluescript 224	1384	99	1384	193	193	447	1	32	33			205
46	HCFMV39	97974 04/04/97 209080 05/29/97	pSportI 56	1603	1	1296	96	96	279	1	29	30			102
47	HCMSX86	97975 04/04/97 209081 05/29/97	Uni-ZAP XR 57	1052	5	786	12	12	280	1	28	29			32
48	HCNAP62	97975 04/04/97 209081 05/29/97	Lambda ZAP II 58	814	1	558	93	93	281	1	22	23			42
49	HCRAF32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR 59	1215	257	1215	356	282	1	19	20				20
50	HCUDC07	97975 04/04/97 209081 05/29/97	ZAP Express 60	478	1	478	147	147	283	1	36	37			69
51	HCWBB42	97975 04/04/97 209081	ZAP Express 61	618	1	618	212	212	284	1	35	36			74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF		
52	HDTAB05	97975 04/04/97 209081 05/29/97	pCMVSPORT 2.0	62	751	1	751	257	257	285	1	21	22	32
53	HE2AV74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	63	780	283	780	433	286	1			16	
54	HE2AY71	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	64	588	21	588	169	169	287	1		16	
55	HE2GS36	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	65	774	272	774	445	445	288	1		37	
56	HE2OF09	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	66	1866	1313	1866	1596	1596	289	1		11	
57	HE6EU50	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	67	1152	117	686	237	237	290	1	20	21	34
58	HE9HUI7	97975 04/04/97	Uni-ZAP XR	68	2483	1577	2448	1620	1620	291	1		14	

169

170

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	NT of AA Signal Pep	5' NT of AA First AA ID NO: Y	AA First AA ID NO: Sig Pep	First Last AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	First AA of ORF
64	HGBA1B	97975 04/04/97 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94	
65	HGBBQ69	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	77	1274	1	1273	105	105	300	1	24	25	43	
66	HGBAJ93	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94	
67	HGFCF08	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	78	1133	4	1042	175	175	301	1	23	24	30	
68	HHFHJ59	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	79	661	1	661	192	192	302	1	29	30	112	
69	HHFHR32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	80	1378	1	1378	358	303	1				13	
70	HHGCN69	97975 04/04/97 209081 05/29/97	Lambda ZAP II	81	1440	298	1440	532	532	304	1	23	24	34	

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
72	HHGDO13	97975 04/04/97 209081 05/29/97	Lambda ZAP II	82 1381	766 1371	993 993	305 1	23 24	24 34	25 34	25 81	25 81
73	HHPFD63	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	83 1706	182 1644	257 257	306 1	24 24	24 25	25 81	25 81	25 81
74	HHSEG23	97976 04/04/97	Uni-ZAP XR	84 573	1 573	160 160	160 307	1 18	18 19	19 71	19 71	19 71
75	HJPAV06	97976 04/04/97	Uni-ZAP XR	85 684	199 684	323 323	323 308	1 1	27 27	28 33	28 33	28 33
76	HKIXL73	97976 04/04/97	pBluescript	86 1036	591 1036	690 690	690 309	1 1	32 32	33 33	33 114	33 114
77	HKMNC43	97976 04/04/97	pBluescript	87 908	1 908	139 139	139 310	1 1	18 18	19 108	19 108	19 108
78	HMEJE31	97976 04/04/97	Lambda ZAP II	88 1102	1 1102	228 228	228 312	1 1	26 26	27 49	27 49	27 49
79	HMSKS35	97976 04/04/97	Uni-ZAP XR	89 1102	1 1102	228 312	228 312	1 1	26 26	27 49	27 49	27 49
80	HNFAE54	97976 04/04/97	Uni-ZAP XR	90 1533	665 1518	347 347	347 313	1 1	26 26	27 293	27 293	27 293
81	HNFIH45	97976 04/04/97	Uni-ZAP XR	91 575	1 575	275 275	275 314	1 1	30 30	31 67	31 67	31 67
82	HNGBT31	97976 04/04/97	Uni-ZAP XR	92 639	1 639	224 224	224 315	1 1	28 28	29 104	29 104	29 104

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Codon	5' NT of AA of Signal Pep	5' NT of AA of Sig Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF	
92	HPFCR13	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	102	1035	602	1035	859	859	325	1	32	33	58
93	HPHAC83	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	103	2218	840	2182	1035	1035	326	1	17	18	17
94	HPMBQ32	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	104	1351	1	1351	18	18	327	1	23	24	86
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	105	2066	51	2052	270	270	328	1	29	30	537
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	226	2057	1	1954	220	220	449	1	29	30	315
96	HRDFB85	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	106	1705	23	1697	233	233	329	1	21	22	201
97	HRGBBR28	97977 04/04/97	Uni-ZAP XR	107	1167	611	1167	53	53	330	1	1	2	263

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Codon	5' NT of AA of Signal Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
98	HSKG N81	97977 04/04/97 209082 05/29/97	pBluescript	108	1907	151	1432	353	353	331	1	23
98	HSKG N81	97977 04/04/97 209082 05/29/97	pBluescript	227	2084	335	2084	537	537	450	1	19
99	HSPA H56	97977 04/04/97 209082 05/29/97	pSport1	109	611	1	576	229	229	332	1	25
100	HE8EU04	209746 04/07/98	Uni-ZAP XR	110	2632	294	2632	337	337	333	1	25
100	HSXBT86	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	228	2143	53	1096	235	235	451	1	9
101	HSXCS62	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	111	2249	1	1953	90	90	334	1	18
102	HTEFU09	97977 04/04/97 209082	Uni-ZAP XR	112	2198	228	2158	400	400	335	1	23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF	
103	HTEKM35	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	113	1043	40	1043	320	320	336	1	20	21	142
104	HTGEP89	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	114	703	1	703	285	285	337	1	29	30	94
105	HTGEW91	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	115	3684	526	1338	584	584	338	1	24	25	37
106	HTOEY16	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	116	1965	127	1915	202	202	339	1	27	28	38
107	HTPCN79	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	117	503	1	503		1	340	1	7	8	70
108	HTSGM54	97977 04/04/97 209082 05/29/97	pBluescript	118	1133	316	1069	423	341	1	12	13	84	
109	HTSHE40	97977 04/04/97	pBluescript	119	1101	118	956	218	218	342	1	31	32	89

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT of AA First SEQ ID NO: Y	AA of Signal Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
110	HTWAF58	97977 04/04/97 209082 05/29/97	Lambda ZAP II	120	282	1	282	137	137	343	1	25	26
111	HTWBHY29	97977 04/04/97 209082 05/29/97	pSportI	121	2635	1593	2489	1654	1654	344	1	25	26
112	HUKRCT1	209007 04/28/97 209083 05/29/97	Lambda ZAP II	122	994	1	932	272	345	1	15	16	221
113	HCE3Q10	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	123	1542	1	1542	143	143	346	1	25	26
114	HCEVVR60	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	124	1390	82	1390	127	127	347	1	32	33
115	HDTAW95	209007 04/28/97 209083 05/29/97	pCMVSPORT 2.0	125	1288	412	1288	571	571	348	1		16
116	HE6EL90	209007	Uni-ZAP XR	126	1517	1	1452	243	243	349	1		9

177

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First A.A of Secreted Portion	Last A.A of ORF
117	HELBUR29	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	127	1073	198	1073	776	350	1			13
118	HERAH36	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	128	300	155	300	202	202	351	1		17
119	HFXBXW82	209007 04/28/97 209083 05/29/97	Lambda ZAP II	129	1275	1	1275	56	56	352	1	23	24
120	HHPTD20	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	130	472	51	472	243	353	1			32
121	HIBED17	209007 04/28/97 209083 05/29/97	Other	131	1950	284	1927	395	395	354	1	72	73
122	HLTER03	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	132	990	1	990	78	78	355	1	22	23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	AA SEQ ID NO: Y	First AA of Signal Pep	5' NT of AA SEQ ID NO: Z	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
123	HOABL56	209007 04/28/97 05/29/97	Uni-ZAP XR 209083 209083 05/29/97	133 1720	565	1720	660	660	356	1	18	19	21	
124	HPMCJ92	209007 04/28/97 209083	Uni-ZAP XR	134	705	28	705	106	106	357	1	28	29	98
125	HPWAZ95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	135	323	1	323	88	88	358	1	27	28	78
126	HRGBR18	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	136	582	1	582	16	359	1	17	18	30	
127	HSUBW09	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	137	1021	1	1021	153	153	360	1	32	33	56
128	HUKCO64	209007 04/28/97 209083 05/29/97	Lambda ZAP II	138	1777	439	1777	521	361	1				2
129	H6EA53	209007 04/28/97 209083	Uni-ZAP XR	139	643	303	643	313	362	1	7	8	31	

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	5' NT of Signal Pep	AA ID NO: Y	First AA of Signal Pep	Last AA of Signal Pep	First AA of Secreted Portion	Last AA of ORF
130	HAGA11	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	140	1220	1	1220		127	363	1	16	17
131	HAGAO39	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	141	721	1	721		415	364	1		14
132	HALSK07	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	142	1468	125	1468	210	210	365	1	29	30
133	HALSQ59	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	143	300	4	300	101	101	366	1	22	23
134	HAIBP89	unknown 05/18/98	Uni-ZAP XR	144	2243	173	2243	311	311	367	1	27	28
134	HBGCB91	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	229	1025	409	1025	624	624	452	1	20	21
135	HBMTD81	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	145	1082	163	1082	357	357	368	1		30

180

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
136	HBXGK12	209008 04/28/97 209084 05/29/97	ZAP Express Vector	146	4313	1153	4313	1313	369	1	18	19	42
137	HFKF107	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	147	1183	1	1183	149	149	370	1	41	42
138	HCQAI40	209008 04/28/97 209084 05/29/97	Lambda ZAP II	148	734	1	734	285	285	371	1	19	
139	HCWHZ24	209008 04/28/97 209084 05/29/97	ZAP Express	149	1405	1	1405	108	108	372	1	34	35
140	HE2GT20	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	150	2890	1178	2890	1178	1178	373	1	31	32
141	HE8EY43	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	151	2399	1181	2399	1265	1265	374	1	30	31
142	HFCEB37	209008 04/28/97 209084	Uni-ZAP XR	152	802	352	802		487	375	1		10

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
143	HFTCT67	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	153	461	24	461	145	145	376	1	37	38
144	HGLAM46	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	154	2388	818	2388	648	648	377	1		18
145	HHGBR15	209008 04/28/97 209084 05/29/97	Lambda ZAP II	155	642	322	642	400	400	378	1		4
146	HJAAU36	209008 04/28/97 209084 05/29/97	pBluescript SK-	156	1251	583	1251	933	379	1	16	17	16
147	HUSIT49	209008 04/28/97 209084 05/29/97	pSportI	157	2127	247	2127	383	383	380	1	47	48
148	HKLABI6	209008 04/28/97 209084 05/29/97	Lambda ZAP II	158	1625	817	1625	1012	1012	381	1	18	19
149	HLMMU76	209008 04/28/97	Lambda ZAP II	159	1687	1307	1687	1296	1296	382	1	28	29

182

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT SEQ ID NO: X		5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT Start Codon	AA of Signal Pep	First AA Y	Last AA Sig Pep	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	First AA of ORF
				Total NT Seq.	Clone Seq.											
150	HMSKQ35	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	160	1842	172	1463	319	319	383	1	30	31	33		
151	HNHED86	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	161	770	1	770	30	30	384	1	31	32	46		
152	HNHEJ88	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	162	519	1	519	242	242	385	1	17	18	24		
153	HNHFQ63	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	163	753	1	753	164	164	386	1	17	18	67		
154	HOECU83	209009 04/28/97	Uni-ZAP XR	164	1400	189	1400	508	387	1	22	23	33			
155	HPTRC15	209009 04/28/97	pBluescript	165	2153	594	2153	611	388	1			13			
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	166	1251	219	1120	389	1							
157	H6EAE26	209009	Uni-ZAP XR	167	882	48	882	155	155	390	1	33	34	153		

183

184

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	179	2911	1103	2858	192	192	402	1	32	33
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	231	1811	20	1811	93	93	454	1	36	37
170	HCUBL62	209010 04/28/97 209085 05/29/97	ZAP Express	180	519	1	519	57	57	403	1	28	29
171	HDSAP81	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	181	968	320	968	476	476	404	1	27	28
172	HE2CT29	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	182	1128	1	1128	111	111	403	1	26	27
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	183	2276	48	2276	88	88	406	1	37	38

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID		5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID of Signal Pep	5' NT of AA of First AA of Signal Pep	First AA of Signal Pep	Last AA of Signal Pep	First AA of Secreted Portion	Last AA of ORF
				No:	X	NT Seq.	Clone Seq.	Start Codon	AA of Signal Pep	NO: Y	AA of Signal Pep	Last AA of Signal Pep	First AA of Secreted Portion	Last AA of ORF
188	HHPSF70	209011 04/28/97	pBluescript	198	951	26	951	162	421	1	16	17	34	
189	HHSAK25	209011 04/28/97	Uni-ZAP XR	199	1740	1390	1740	1534	1534	422	1	19	20	31
190	HIASB53	209011 04/28/97	pBluescript	200	1707	401	1195	652	652	423	1	26	27	126
191	HJABZ65	209011 04/28/97	pBluescript SK-	201	779	1	779	23	23	424	1	26	27	68
192	HJPBB39	209011 04/28/97	Uni-ZAP XR	202	1617	188	1605	182	182	425	1	28	29	91
193	HLHSK94	209011 04/28/97	pBluescript	203	1974	1	1794	112	112	426	1	26	27	379
194	HLHTC70	209011 04/28/97	pBluescript	204	1057	229	1057	365	365	427	1	23	24	22
195	HLMIW92	209011 04/28/97	Lambda ZAP II	205	721	1	721	244	244	428	1	25	26	46
196	HLTCY93	209011 04/28/97	Uni-ZAP XR	206	2465	988	2465	1225	1225	429	1			4
197	HLTDB65	209011 04/28/97	Uni-ZAP XR	207	1480	1	1480	371	430	1	15	16	143	
198	HMSHM43	209011 04/28/97	Uni-ZAP XR	208	872	1	872	35	35	431	1	18	19	36
199	HMSHQ24	209011 04/28/97	Uni-ZAP XR	209	1779	16	1779	148	148	432	1	24	25	36
200	HNFAH08	209011 04/28/97	Uni-ZAP XR	210	2110	592	2110	611	611	433	1	18	19	191

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO:	NT Seq.	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of Clone Seq.	AA SEQ ID NO:	First AA of Signal Pep Y	Last AA of Signal Pep	First AA of Secreted ORF	Last AA of Secreted ORF
201	HNGAO10	209011 04/28/97	Uni-ZAP XR Vector	211	938	1	938	107	107	434	1	27
202	HNGBE45	209011 04/28/97	Uni-ZAP XR	212	1551	1	1551	114	114	435	1	21
203	HNHAZ16	209011 04/28/97	Uni-ZAP XR	213	997	1	997	202	202	436	1	24
204	HNHCM59	209011 04/28/97	Uni-ZAP XR	214	1496	1	1132		165	437	1	28
205	HOSFM22	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	215	1308	501	1308		809	438	1	1
206	HPHAC88	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	216	1705	384	1705	549	549	439	1	23
207	HCDE095	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	217	999	608	999	273	273	440	1	22

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The 5 overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain 10 multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT 15 of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified 20 as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted 25 first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and 30 otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic 35 methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

- 5 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).
Polypeptides of the invention also can be purified from natural or recombinant sources
10 using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

- Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

- In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

- As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other 20 words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF 25 (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between 30 a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result 35 of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:
Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,

- 5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be

- 10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
15 subject sequences are either both nucleotide sequences or both amino acid sequences.

The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.

- 25 For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired 5 residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. 10 This time the deletions are internal deletions so there are no residues at the N- or C- termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query 15 sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or 20 activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in 25 the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. 30 Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be 35 deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological 5 activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1 α . They used random mutagenesis to generate over 3,500 individual IL-1 α mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible 10 amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

15 Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form 20 are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show 25 substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main 30 strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions 35 where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham 5 and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the 10 protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues 15 Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, 20 where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino 25 acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins 30 with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 5 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers 10 as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-15 450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the 20 deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the 25 deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 30 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

35 Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any 5 combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in 10 the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue 15 identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. 25 Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active 30 fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having 35 antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 5 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at 10 least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to 15 methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if 20 it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is 25 meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, 30 as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion 35 proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the

polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5 Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of
10 the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final
15 preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins
20 facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG)
25 can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules
30 together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the
35 fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

- Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In 5 preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.
- 10 Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

15 Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the 20 latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then 25 transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The 30 expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

35 As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila S2* and *Spodoptera Sf9* cells; animal cells such as CHO, COS, 5 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, 10 pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium 15 phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

20 A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most 25 preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also 30 be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial 35 modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

5

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

10

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

15

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing

20

the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

25

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence *in situ* hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

30

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for 5 contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers 10 for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The 15 following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-20 3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and 25 technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of protein include those detectable by X-radiography, NMR or ESR. For X-30 radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

35 A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human
5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The
10 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves
(a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene
15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to
20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired
25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such
30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a
35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and 5 polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

10 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells 15 from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

20 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic 25 cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency 30 (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a 35 polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in 5 treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, 10 differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, 15 glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune 20 inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to 25 treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The 30 administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may 35 inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or

5 IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

- 5 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae,
- 10 Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g.,
- 15 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps,
- 20 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

- 25 Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Nocardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter,
- 30 Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis,
- 35 and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

- 5 Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

10 Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

- 15 These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or

20 diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (*ex vivo* therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase 5 regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue 10 regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate 15 nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, 20 Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

25 A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular 30 trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body.

35 For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

5

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

10 (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural 15 or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

20 Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed 25 polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results 30 in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule 35 activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with 15 a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or 20 decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic 25 surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian 30 rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a 35 food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

- 5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined
10 from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

- 20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

- 25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

- 35 The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with 5 the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with 10 the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in 15 Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as 20 defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at 25 least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method 30 comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino 35 acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone

5 identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

10 Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as

15 defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide

20 molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition

25 associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a

30 sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

- Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.
- 5
- Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.
- 10
- Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.
- 15
- Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.
- 20
- Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.
- 25
- 30
- 35

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase 5 the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

10

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. 15 Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector 20 "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
	Zap Express	pBK
25	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSport 2.0	pCMVSport 2.0
	pCMVSport 3.0	pCMVSport 3.0
	pCR®2.1	pCR®2.1
30	Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are 35 commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1	

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation 5 of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain 10 DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed 15 into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

20 The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises 25 a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited 30 sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized 35 using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P - γ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate.

5 These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

10 Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with

15 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product

20 is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

25 Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

30 Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to

35 generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then 5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA 10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

15

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR 20 using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, 25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is 30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are 35 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This 5 primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on 10 either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

15 A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product 20 into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

25 The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are 30 identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The 35 cells are grown to an optical density 600 (O.D.^{.600}) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by 5 centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrolo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high 10 affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 15 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 20 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes 25 an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and 30 XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

- 5 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed
10 in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified,
all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell
culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at
15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit
15 weight of cell paste and the amount of purified protein required, an appropriate amount
of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50
mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a
high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer
20 (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is
then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by
centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M
NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine
25 hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the
pellet is discarded and the polypeptide containing supernatant is incubated at 4°C
overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles,
the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20
30 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by
vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing
for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential
filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a
5 stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion
10 (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280}
15 monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded.
20 The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient
25 polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that
30 express the cloned polynucleotide.
35

Many other baclovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baclovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baclovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baclovirus DNA ("BaculoGoldTM baclovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGoldTM virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

- 5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture
10 and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in
15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

- To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection
20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins
25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

30 **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved
5 with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include,
for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),
10 pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109),
pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used
include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1,
Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO)
cells.

15 Alternatively, the polypeptide can be expressed in stable cell lines containing the
polynucleotide integrated into a chromosome. The co-transfection with a selectable
marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation
of the transfected cells.

The transfected gene can also be amplified to express large amounts of the
20 encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing
cell lines that carry several hundred or even several thousand copies of the gene of
interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin,
J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and
Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is
25 the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991);
Bebbington et al., BioTechnology 10:169-175 (1992). Using these markers, the
mammalian cells are grown in selective medium and the cells with the highest resistance
are selected. These cell lines contain the amplified gene(s) integrated into a
chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the
30 production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the
expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession
No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et
al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the
35 CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g.,
with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the
cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by 5 procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a 10 heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

15 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for 20 transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are 25 trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of 30 methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose

- 5 binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the
10 activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in
15 Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

- 20 For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that
25 the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a
30 heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAAACACACATGCCACCGTGCC

CAGCACCTGAATTGAGGGTGCACCGTCAGTCTTCCTCTCCCCCAAAACC

- 35 CAAGGACACCCCTCATGATCTCCGGACTCCTGAGGTACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC

AGCACGTACCGTGTGGTCAGCGCCTCACCGTCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCAAACCCCC
ATCGAGAAAACCATCTCCAAAGCAAAGGGCAGCCCCGAGAACCAACAGGT
5 GTACACCCCTGCCCTATCCCAGGGATGAGCTGACCAAGAACCAAGGTAGCCT
GACCTGCCTGGTCAAAGGTTCTATCCAAGCGACATGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCGTGCTGG
ACTCCGACGGCTCCTCTCCTACAGCAAGCTACCGTGGACAAGAGCA
GGTGGCAGCAGGGAACGTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACACTACACGCAGAAGAGCCTCTCCGTCTCCGGTAAATGAGTGC
10 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell
25 Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at
30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

35 The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art.

(See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulian et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

30

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The

- 5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x

- 10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in

- 15 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of

20 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off 25 PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl₂ (anhyd); 0.00130 mg/L

- 30 CuSO₄·5H₂O; 0.050 mg/L of Fe(NO₃)₃·9H₂O; 0.417 mg/L of FeSO₄·7H₂O; 311.80 mg/L of Kcl; 28.64 mg/L of MgCl₂; 48.84 mg/L of MgSO₄; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO₃; 62.50 mg/L of NaH₂PO₄·H₂O; 71.02 mg/L of Na₂HPO₄; .4320 mg/L of ZnSO₄·7H₂O; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L-Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site “GAS” elements or interferon-sensitive 5 responsive element (“ISRE”), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or “STATs.” There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is 10 Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

15 The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase (“Jaks”) family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

20 The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and 25 (b) Class 2 includes IFN- α , IFN- γ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

30 Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are 35 known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u>	<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>	<u>STATs</u>	<u>GAS(elements) or ISRE</u>
	<u>IFN family</u>							
5	IFN- α /B	+	+	-	-	-	1,2,3	ISRE
	IFN-g		+	+	-	-	1	GAS (IRF1>Lys6>IFP)
	IL-10	+	?	?	-	-	1,3	
	<u>gp130 family</u>							
10	IL-6 (Pleiotropic)	+	+	+	?	?	1,3	GAS (IRF1>Lys6>IFP)
	IL-11 (Pleiotropic)	?	+	?	?	?	1,3	
	OnM (Pleiotropic)	?	+	+	?	?	1,3	
	LIF (Pleiotropic)	?	+	+	?	?	1,3	
15	CNTF (Pleiotropic)	-/+	+	+	?	?	1,3	
	G-CSF (Pleiotropic)	?	+	?	?	?	1,3	
	IL-12 (Pleiotropic)	+	-	+	+	+	1,3	
	<u>g-C family</u>							
20	IL-2 (lymphocytes)	-	+	-	+	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	+	6	GAS (IRF1 = IFP >> Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	?	6	GAS
25	IL-15	?	+	?	?	+	5	GAS
	<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	-	5	GAS
	<u>Growth hormone family</u>							
	GH	?	-	+	-	-	5	
	PRL	?	+/-	+	-	-	1,3,5	
35	EPO	?	-	+	-	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
	<u>Receptor Tyrosine Kinases</u>							
	EGF	?	+	+	-	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	-	1,3	
40	CSF-1	?	+	+	-	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., *Immunity* 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCG
10 AAATGATTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCGAAATG
20 ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCC GCC
CTAACTCCGCCATCCGCCCTAACTCCGCCAGTCCGCCATTCTCCGC
CCCATGGCTGACTAATTTTTATTATGCAGAGGCCGAGGCCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCCTTTGGAGGCCTAGGCTT
TGCAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, 30 alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using Sall and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HE LA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and 25 Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 30 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to 35 generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final 5 concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

20 After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

25 The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jak-STAT signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfet U937 cells with the GAS/SEAP/Neo construct produced 10 in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM Na₂HPO₄.7H₂O, 1 mM MgCl₂, and 675 uM CaCl₂. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then 20 resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well 25 plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. 30 Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, 5 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or 10 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

15 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

20 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

25 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) 30 containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

35 Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 μ g/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 μ l of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 μ l supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ μ l of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

- To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTCCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GC GG C CT CG AG G G G ACT TT C CC G G G G ACT TT C C G G G AC
TT C CAT CCT GCC AT CT CA ATT AG:3' (SEQ ID NO:9)
- The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
5':GC GG CA AG CT T T T G C A A A G C C T A G G C:3' (SEQ ID NO:4)
- PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)
- Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTCCCGGGGACTTCCGGGGACTTCCGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATACTCCGCCCTAACTCCGCCA
20 TCCCGCCCTAACTCCGCCAGTTCCGCCATTCTCCGCCCATGGCTGACT
AATTTTTTATTTATGCAGAGGCCGAGGCCGCTGGCCTTGAGCTATT
CAGAAGTAGTGAGGAGGCTTTTGAGGCCTAGGCTTTGCAAAAAGCTT:
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the 10 following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven 15 heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

20 Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

25

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

- 5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100
10 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

- 15 To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

20

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

- The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

- 30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

5 Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodynne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr 10 with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of 15 alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodynne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of 20 Loprodynne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ 25 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum 30 manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

35 Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 5 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
10 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

- 15 Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-
20 POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

25 Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

- As a potential alternative and/or compliment to the assay of protein tyrosine
30 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
35 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then 5 rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C 10 until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyn filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts 15 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and 20 Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

25

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from 30 these RNA samples using protocols known in the art. (See; Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

35 PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTHERM Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and
5 Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated
10 according to Example 2 are nick-translated with digoxigeninideoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

15 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv.
20 et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated
25 disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample,
30 and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a
35 sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

5 The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

10 Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

15 Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

20 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

25 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g}/\text{kg}/\text{day}$ to 10 $\text{mg}/\text{kg}/\text{day}$ of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 $\text{mg}/\text{kg}/\text{day}$, and most preferably for humans between about 0.01 and 1 $\text{mg}/\text{kg}/\text{day}$ for the hormone. If 30 given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g}/\text{kg}/\text{hour}$ to about 50 $\mu\text{g}/\text{kg}/\text{hour}$, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending 35 on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes 5 of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. 10 Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric 15 acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; 20 U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is 25 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are 30 known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood 35 of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as 5 ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, 10 manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of 15 about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed 20 into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials 25 are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical 30 compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

- 5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 20 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is 30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

10 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to 15 transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

20 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

25 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is 30 required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made 5 on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel 10 clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from 15 different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be used to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly 20 described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent 25 applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

Sequence Listing

(1) GENERAL INFORMATION:

5 (i) APPLICANT: Human Genome Sciences, Inc., et al.

(ii) TITLE OF INVENTION: 207 Human Secreted Proteins

10 (iii) NUMBER OF SEQUENCES: 800

(iv) CORRESPONDENCE ADDRESS:

15 (A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

20 (C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

25 (F) ZIP: 20850

30 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

35 (B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

40

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

45 (B) FILING DATE:

(C) CLASSIFICATION:

50

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

55 (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

- 5 (A) NAME: Kenley K. Hoover
(B) REGISTRATION NUMBER: 40,302
(C) REFERENCE/DOCKET NUMBER: PZ007PCT

10

(vi) TELECOMMUNICATION INFORMATION:

- 15 (A) TELEPHONE: (301) 309-8504
(B) TELEFAX: (301) 309-8439

20

(2) INFORMATION FOR SEQ ID NO: 1:

- 25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

30	GGGATCCGGGA GCCCCAAATCT TCTGACAAAAA CTCACACATG CCCACCGTGC CCAGCACCTG	60
	AATTGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCCC AAA ACCCAAGGAC ACCCTCATGA	120
35	TCTCCCCGAC TCCTGAGGTC ACATGGGTGG TGGTGGACGT AAGCCACGAA GACCCCTGAGG	180
	TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCCGGGG	240
40	AGGAGCAGTA CAACAGCACG TACCGTGTGG TCACCGTCT CACCGTCTG CACCAGGACT	300
	GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCC ACCCCCATCG	360
	AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCC	420
45	CATCCCGGGA TGAGCTGACC AAGAACCAAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT	480
	ATCCAAGCGA CATGCCGTG GAGTGGGAGA GCAATGGCA GCCGGAGAAC AACTACAAGA	540
50	CCACGCCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCT CTACAGCAAG CTCACCGTGG	600
	ACAAGAGCAG GTGGCAGCAG GGGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC	660
	ACAACCACTA CACCGAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC	720
55	GACTCTAGAG GAT	733

60 (2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

10 Trp Ser Xaa Trp Ser
 1 5

15 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

25 GCCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTCCCCG AAATGATTTC 60
 CCCGAAATAT CTGCCATCTC AATTAG 86

30

(2) INFORMATION FOR SEQ ID NO: 4:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGGCAAGCT TTTTGCAAAG CCTAGGC

27

45

(2) INFORMATION FOR SEQ ID NO: 5:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCCCGAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTCCCCG
 AAATATCTGC CATCTCAATT AGTCAGCAAC CATACTCCG CCCCTAACTC CGCCCATCCC

60

120

60

268

GCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCAT GGCTGACTAA TTTTTTTTAT 180
TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCCT 240
5 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10 (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

20 GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

25 (2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

40 (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
45 (B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

50 GGGGACTTTC CC 12

55 (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
60 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 GGGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60
 CCATCTCAAT TAG 73

10

(2) INFORMATION FOR SEQ ID NO: 10:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 256 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTCGCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60
 25 CAATTAGTCA GCAACCATAAG TCCCGCCCCCT AACTCCGCC ATCCCGCCCC TAACTCCGCC 120
 CAGTTCCGCC CAATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180
 30 GGCGGCCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTG GAGGCCTAGG 240
 CTTTTGCAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2526 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

45 GACAGGCTAT CGGAGAACATCT GAGAGCTGGG CCGGGCAATT CCTCCAGYTA CCCCTTGAC 60
 CTAAGTCCAG TCACACATTT CCCAAAGTTT CTCCTTGTC TAACCCTGGT CTGGCTGGTT 120
 50 TTGRGGRCTT GAGAACGGGT CAGGGACTCC AGGCCAAGTC CAACAGAGAC CCCAAACCCA 180
 CCACACACCA GCAGCCACAA CCTCACCAACC AACAAAGAGG ACTTTTGAGG GGCCACAAGT 240
 AAGAGGTCAAT TTCTGGAATG GACTCAGACC TTTAACAGG AGAGTTGAGC ACTTCCAGKS 300
 55 AGTTTTTAAG CAAGGCATGG GGAACAGGGGA ATAGAACCTT TCAAAGAGGT TGCCCAGAGA 360
 AAAGCTGGGC CTCTTGCATT CGGCTTCCTT GGAGCAGCCT CTTCTGGCAG AAAGCCATCA 420
 60 GGTGCTCAAT CATCTCTCC TGGCCAAGGC TCTGACCATG CTTAGTACTG GAATAGAGGT 480

	GGCCAGGCC CCAGCGACTC TTCTTGGCCT GATGTTGTC CTCACAGGCA TGCCACGTGG	540
5	CCTGAGATGA TTCAGAACAA ATCATGCTAA CTTTGAATCC ATCCAGCCAC TTGCAAATGA	600
	TAATCAGAAG TCAGCTTGT CACTGTTAGA AAGAAACTAA CAAAAGAGAA CCCAGAGCAA	660
	TCTAGAACATCT TTGAGTGCTT GGCTTCCAA GGATACTGCG GAGACTCTGG CCAAGCTGAT	720
10	GAMCTTCTGA ARTGTCACTG GCACCATATG CAACAAGAAC CACCATTAC TGAGTAGCTA	780
	ATGGGTTTGG GCCCTGGGAC ATTCCATCTG AGGTCCCTCC TGAACATGTC ACTCCACAGC	840
15	AGAGGACCGG TTGCAGCTTA CCCAGAACCA CTCCCTCCAGG AGAGCTGGAT GTTTTGGTGTG	900
	CAACACCTTG AGCACTGACT GCTATTGTTA AAAAAAAAGCC TTTGCTGCAT TCGGAGGACT	960
	GCCCCGTGCC CTGAGGTGAC TTCCCTAACTA TGTGGTTCA TTAGCGAATT TATTTTTGT	1020
20	GCTGGGTGGA CATTGTATT TTGTTAGGTT GCTGTTAAG CTCAAGTTTG CTGTGCTCTC	1080
	TGCAGCTACA AAACATCTTG GCATATTAA GAKTGGCTTT TATAAATAGC TTTATTCTGA	1140
	TATTAATCAG ATTCCCAACT TTACTGAGAA TTAAGGACTG GGGTACTTTA AAGAAATGCA	1200
25	AATAGCAATT GAAGAACAC TGCTGCAGGT GGTAGCCCTG GCTAGACTGA ATTACACTAG	1260
	AAATCAGCCA GAAGGAAGCG TCCCTGGGAT CCCAGATCAC TCTTTTTTT TTTTTTTITA	1320
30	AAAGGGCAG CCCCTTGATG GCTCATCTCT CTGAATAACA GTTACGTCTT CATATCGATA	1380
	CCAGATGCCT TCTTCATCAT GCCACTGAAG CCACTCACCA CCTTCAGAAC CATGCCAAC	1440
	TCTGTCAGAT TCACCTACCC ACAAACAAAGG AGGCACGTTT GGCACAAAGT GTTGTCTCC	1500
35	AGGTCCAAGT GGACTCTACA GAGTGCTTGA CCTCAACACA CTGGATTCCA GGTGGACTGG	1560
	ACCAAGAGCA GGCAAGACA CGGGAACTGA AAAACTCCAC AGGTTTGGA GAATAGAAAT	1620
40	GAAAAGCCAC GTCATATAAC TCAAGAATAA ATGGTGTTTT GGAAATTITA AAATTATCAT	1680
	CGAAGGTGGT GAAACTATT CAGGCCAAA TGAAAGGAAA TCGCCAGTTG GGGATGAAAT	1740
	CACAGAGCCT GTGTTTATG ATATGGTTGG ATGTCCACTG ATGAAATTAA AAAGGAGTTT	1800
45	CATTTTTAAA AGTGCCTCATG ATTCTACATA TGAGAATTCTT TTAGGCCAAG AAACGTGCT	1860
	TGGCTCAGAG GTGTTGGAA TAAAGCAGA GAGAAGCCAT TCGTGTGCT TAGAACCAAG	1920
50	GATGGTCATG TACACAAAGA CCATCGAGAC GGCCATTCTT GTTACAAAA CACTTACCAA	1980
	GAAAGCACTT TGTAGGGAA CTTTAGTAAG TTCTTCAT TTCAATTATGT TTCTTCCAAG	2040
	GAAACAGGAG AGACTGAATT AATAATTCTC TCTTCTCT TAAGCACTTT TAAAATAATA	2100
55	AAGTACATCT TGAAATTGGG GGGGCATCT CTGATTAAA AAAAGAAAAA GGCTGCTTGA	2160
	TGTATGTTAT GCAGAGACAC TCTGCCTCTG GTGGCTGCAG AGCAATACCC AAGCCTCATT	2220
60	TGGAAGGCTC AACATTGGAA ATTGCACTTT AATTGATTAA TCCTCAATTG ATGTGGCCTT	2280

ACGGGATGGT GGGTCTGGGA CCCCAATTCA TTCTTATCTG CCAAAGAATT ATCTAGAAC	2340
5 ACATCAAATA CCAGCACCCC ACCTGCACAA TGGGGGTGGA AAACTTTGT ATCCCTAAGC	2400
ATATTATTT ATAGTGTCTG CCATGCCATG TGAAATACT TTATTTTAA CCTCAGGATT	2460
TAAATAAAGT AAACACTATG ACATTTAAAA AAAAAAAA AAAACTCGAG GGGGGCCCGG	2520
10 TACCA	2526

15 (2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 CACTGCACCA GCTTTGTAT CTGAAAATG ATGATAATAC CAACACCTTC TTCTTGGGT	60
ACTGAAGATG AGAGAACATG ATATGTGTAAGTGCCTTCC ACAATACCCA GAACATAGCA	120
AACATGTAAT GAATGTAGTA ATAGTAATTA TTTTATTTTC TTTTGATTCA GTTGGGACTA	180
30 TGTCAGCTG TAACAGAATA CCCAAAATAA CTGTTTAAA CAAATTAAAG TTTWGTGTC	240
AAGTTTGTGTT ACGAATTCAAG ACAATCCAGG GCTTTTATAG ATGCACCAGG ATCAGCAGGT	300
35 ACAAAAGGCAT CTTTCTGAT TTCTGCCAGT CTCAATGCAT GGTTGCAAT CCAGARTCCA	360
RGATGGCAGT TCCAGCCCTG GTTACGCCCA TATTAGCACA CAGAAAGAAA GAGAAAGGGA	420
TGTGCCTCTT CACTTTAACATC ATAGCTCCCCA CTAGATGCAC CCACACTTC TGCTGATACT	480
40 CCATTAGCTA ATGCTTGCTT ACATGGTCAC ACTTAGTTTC CAGAGAGACA TGTCTGGACA	540
GTCATGTGCT CAATTAATAT CCAAGTGTCC AATTACTGAG AAAAAAGAA ACTAGCACCT	600
45 TTGCTTGGTT GCATTCCTCT TAGCATAAGC CACATTCTT TTATGAAGTT GTCTCAGTT	660
ACTTGGATGC CTCAGTTGTC CTTCAWTTA GAAAWCYCC TKGGACAYCC TGAAWCTGAC	720
50 TTCTTTGTC ATCAGCACCA TCACTACCAC TGCCYTCTTC AAAGCCACCA CGTTCTGTCC	780
CCAGGATGGT TGCAACAACC ACCATAGGGA CTTTTGCCT TCTACTTCCA CACAATAGNC	840
CAGAGTAAGC TTTTGAAAAT GTAGGTAGA TCATGTCTCT CTCTTCCCT TCAAAACCC	900
55 CCCGATGGCT TTTCATATTA CTCAAAAGAA AACCTAAAAC TTTGCTGTGA GATCTATGTG	960
ACCCGGCTTA TTCTTCCCTCT TACTTTATCT CTGTATTGCT CTTCCTCACT CTACTCCAGC	1020
60 CATCCCCACCT CCTTGCTGCT TGTCTTATAC TCCTAAAAGA AGTTCACTCT TCCCTTATGA	1080

TATTGCACT TAAAATAGAA AAAAAAAA AAAAAAAACT CGAGGGGGC C 1131

5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

15	GGCACGAGTA GCATTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTATGT	60
	GATGTCCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA	120
20	GGCTGGAGAG ATCATATTT TGTTATTAAA CTGGAGCTTC TCCATCCTTC ACATTGTTGA	180
	TGTCCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCGCGTAGCG GTTTGAGCCA	240
25	GAGAATGACA GCTCTGGTTT GGAGAAAAGG CCCGGATGGT GGCTCTAGAA AGCCCATCCT	300
	TCTGCTCTTC TTTTTCTCC CCCTTATATT GTGCTTCAT TCATTCAATT ATTCAATCAA	360
	CATTGTTGA GCACCTAATA TGTGTCAAGC TCTGTGCTAG CCTCTGGAAA ACCTGCCCTC	420
30	ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGGTTGTCA	480
	GGGTCTCACA GAGCAGTGGC CCCTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC	540
35	GAGGATGGTG TCAGAGCTAA CTGAAGAACG AGAGGGAGCT GCACCASCAG GGGTTGGAAC	600
	TGAAGGTGGC AGTGCCTGGA GTCTTGATTC CAGCAGAGGG AGAGCAGTCT GTGAAAAGGC	660
	ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACCTCA CGTAGTTCTG GATGGCSCTG	720
40	GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA	780
	TCCCAATAAA CCCATTGGAA ACCAAAAATT TAAGTCAGAA GTGCATTAA GGCTGGTCG	840
45	AGTACAATGA TTTTTACAAC GAATTGATCA CAACCAGTTA CAGATGTCTT TGTTCCCTCT	900
	CCACTCCCAC TGCTTCACCT GACTAGCCTT TAAAAAAAAA A 941	

50

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 843 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

60

	CNAGGGATAA CCCCAAAGNT GGGAAATAAA CCCTCAATT AAGGGGAAC CAAAAGCTG	60
	GGAAGTTCCC CCCCAGGGTG CGGGCCNGNT CTAGGAACTA GTGGAATCCC CGGGGGCTGC	120
5	AGGGAATTCTG GCACGGAGTG GGAATGTTGT TTGTATGATA CTATTCCAC AAATGCATT	180
	GAGACTTGGT KTGTGGCCTA GGACATGGTC AATTCTTTYT AAATATTCCG TGAATTCTT	240
10	TAGTGCATAT TCTCCGATGG GGGCTGTGGG GACAGAGTTC TAAATATGCC CATTAGATTA	300
	AATCTCTTCA TTCTGTTGCT CACATCTTCT ATATCCTTAT TAATCTGTCA ATCTCTTCAA	360
	GAGAGGTGTT ATTAAAATCT CTCACTGTAT GTGTCACTTT GCCCTTAAAA TTCTGATGAT	420
15	TTCGTTTATA AATGGTTATA ACCATTTCC AGGAAGAACAA TTAAAGAACT TTCCATTGGC	480
	ATTATCCAGT TTCCCTCAAA ATACTGGTTT TTTTTATTTT GGCTNCTAAG CAGCTATGAA	540
20	TCCAGTTCT CAGAACCCCT TGTCTCAAGG CATTGTTTC CAGATTACCT TGTTAGCATC	600
	CACACTATGG GCTATTTAG AAAAACAAAA AAAGTATCAA AATCATATAG CTATGATTTT	660
	CCTGTGCTTG AAGGAGCCTT AAAGCTCATC TAGTCCAGCC AGTATTTGTT CATCCAAATT	720
25	CTGCCAAGAA ATCTCTATTG TCAAGATATT CTTTACCATC TTGGGACAT TCTCATTATT	780
	AGAAACAAAT CCTAAGAAGA AATTCTGCCA TAKACAACCC ATCCGTTCTT TAAAAAAA	840
30	AAA	843

(2) INFORMATION FOR SEQ ID NO: 15:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1018 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45	CTGTAATTAA TAATTTCAT ATACCGTGCT TTGATTCTAA TTTTATTTTG TGAGTTCTCT	60
	GAAGGTTACA TATACAGAGT GCTTCAGGAA TGATCATTTT GTTATTATTTC ATGCTTCTTA	120
	ACAATGTTGT TTTAGTCCAA GAAGATAATT GCCAGAGAAA GAATACAGTG CAGGAAAGAA	180
50	GARGCTGGAG CCAGTGGTGA AGARGGATTG AGARGACAGA CATTGIGGGA ATGAAATCAT	240
	GAATAATCGT GTTTTGAAT TGTCCAAAAA CTTCTACAAA CCATGAAATG TTGGAGTTA	300
55	AATCTAATTG TTGAAAATT CCCCACATT CTTGTATCCC TTAGGTTGAG CATAATTCCA	360
	CATCCGTGGA CTGATGCACT TCCCAAGAGG GGGCCTCATT AACTCTCCG AGGCAGCAGC	420
	AGCAAGGGCA CCCCTCCCTT TCCCCCACA CCCCACTTCT CATGGCTCTT CTTCTCTCA	480
60	TCTCATGCTT AGGTTAGAAA AGGGCACAAG GTAAGGAAGC CCTTGGAAT AGGCTGAATC	540

TGGCTATCTA ATTTGGTGCC AAATACCTAA TGTGCTTGAA TTTAAAAACA GCAAACATGT	600
AGAAAAGTAA TTATAATTAT GAGGCCAGTT CTTTAAGCTA GCTTTTTTC CCCTCTCAA	660
5 CAGCATATTG GCTTGGATGT CAGCAGGAGA AAGTGTGTT TGCAATACAC ATAATGCATA	720
TATGGTCCTG TTAGCAATCT ATAGAAAATA GATATTGCTC ATTAAGGTAA ATATTTTGT	780
10 TGATGAATGA TCTGGAATGG TCTGGACTTG TTGTGTGAAC AGGAAATTGC TCTGTAGGCT	840
TTGACTTGTG AGGTAAAGAG TGAGGCTGGT AAGATTAATT AAAGTAAATA CTGTGACAAT	900
15 AGGATGTCAA AACCAAAAAC GTGTTCTGA AACTCAAGGA ATTAATGACA CATAGGGAAG	960
TTTTGCCAT ATTAACCATA GAGTAGGAGA GGCAAGTCAA GAATAAAAAA AAAAAAAA	1018

20

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 661 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TTTAAGAAAT TAGTGAATCC CCGGNIGCAG GGAATTGGC ACGAGGAGGA GGCGGTCA	60
TGGCAGGAGC GCAGGGATGGC AGCTGYTCCC CCGGGGTTGCA CCCCCCCAGY TCTGCTGGAC	120
25 ATAAGYTGGT TAACAGAGAG CCTGGGAGCT GGGCAGCTG TACCTGTGGA GTGCCGGCAC	180
CGCCTGGAGG TGGCTGGCC AAGGAAGGGG CCTCTGAGCC CAGCATGGAT GCCTGCCTAT	240
40 GCCTGCCAGC GCCCTACGCC CCTCACACAC CACAACACTG GCCTMTCCGA GCTGCTGGAG	300
CATGGAGTGT GTGAGGAGGT GGAGAGAGTT CGGCGCTCAG AGAGGTACCA GACCATGAAG	360
GTGCGCAGGG CAGGGCTCGG ACCTACCCCA GGAATGTCCT GCCCTGGAA TGACAACACA	420
45 GTCCACACCA TGCACGGGAA GGCAAACAGG GGCAGCTGAC CCAGCCCAGG GGTCAGANGA	480
GGTCTTGGCG AGGAAGTGGC AGCTAAGCTG ATACCTGATA TGCACWAGKC AGCCARGYGG	540
AGACAGGCAA GGAAGAAGCT TGTTTGAGG ACAGAATTCT CTAGATCACT CAGCACCAC	600
50 TGGCTTTTGG GGCTTTTGT TTTATTTTGT TTTTGAGACG GGGTCTCGCT CTGTCGCCA	660
N	661

55

(2) INFORMATION FOR SEQ ID NO: 17:

- 60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 553 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGCACAGGGC TATTTGCCCA TCTCTCCACA TGACAGAACT GCTCTAAGTT TCTTTGCTGC	60
10 TCTTCTCAGC TGTCAAGACGG CTTGCTGCTT GTTTTCCACA CCACCATGTC TATTCTTGC	120
TGTCCCTTWAC TCTGCCTGTT TTTTTCCCTT TGTATTTCTT CTGGCTCTTG TCCCCTTTCC	180
15 CACGTGTcWC AGCTTTCCCTT TATTGCCACT TTCAGTCAGA GCAGTCCTGT GCTTCTGGTG	240
CCGGCATAACA ATACTTACTT GAGTTTCTTG GCTTTTCTTG ACTGTGCATC TCTTACTTCA	300
ACATAGGAAT AGCCTGTcAT AGAATTCTC CAGTTCCAGG CCTCAAGAGG GAGAGTGCCA	360
20 GAAAATTGAG ACTGTTTCC CTGCTTGGA TTGAATTcAT AAAGCAAAAC CAGTGTGTTGT	420
GTGAGGGTTT GCTGTGTcAT GCCTATAGGT TGTTGGGTG CAAACCTATA GAATCCAGCC	480
25 TGCGAAAAGA AAGRAACCAG AGAATANCAG CATCAGAACCA ATGCTTGACA TCATTTCTCA	540
ATCAAGCAGT CCA	553

30

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 869 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

40 GGCACGAGCT GCCAACACTG AGGTCTTCGT GGCTTCTCAC ATCTAGATGT ATCCCTCTCA	60
AATCTATCCT CTATCCAGGC ACCAGATTGA GGTATCTAAA ATGTCACATT TCCAGTTACT	120
45 CCTTCTTATA CTAGCCCCAT CAACTTACAA GATAAAGTCC AAGCCCCCTTC ATATGACAAA	180
CCACACCCCTG CTTAACCTCTC CAGGTTTGAA TCCTTCATCT CCTACTTTAA ACTTTAAAAC	240
50 CCAGCAGCAC GAAAGTGTCT CCTATGCATG TTGCCATATG CGTCTCTCC ATCATGCATT	300
TGCCTGAGCA AGATGTCTTG AGTTAACATC TTATTCTTTA AGACTCATIG TGGTGGTAGA	360
CAGCCTTTAA TAACGGATCC TTGGCCAGGC ACAGTGACTC ACACCTGTAA TCCCAGAACT	420
55 TTGAAAGGCC AAAGAAGGAA GAAAGCTTGA GGCCAGTAGT TTGAGACCAG CCTGGGAAAC	480
AGAGAGATAT CCCATCTGTA CCAAAATTT AAAAATAT TAGCAGGGAG TAGTGGCATG	540
60 CACAAGTGGT CCCAGCTCCA TGGGAGASTG AGGTAGGAAC ATCACTTGAG CCCAGGAAGT	600

CAAGGCTGCA GTGAACCATG ATCAGAACAT TGCANTCCAG CTGGGTAAAC AGAGTGAGAC 660
 CTTAGGTCAG AAAATGAAT AAATAAGCAT AAAATTAA AACCTAGCC AGGCATGGTG 720
 5 GCACACATCT GTGGTCCCTG CTACTTAGGA GGCTGAGGTG AGAGGATCCT TGAGCCCAGG 780
 AGGTCAACAC TACAGTGAGC TATGATTGTG CCACTAAACT CCAACCTGGG TGAAAAAGCA 840
 AAACCTGCC AAAAAAAA AAAAAAAT 869
 10

(2) INFORMATION FOR SEQ ID NO: 19:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 959 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

25 GCGGAGCCGA GATCGTGCCA TTGCACTCCA GCCTGGGCAA CAAGAGTGAA ACTCTGTCTC 60
 AAAAAAAA AATTATAATA CTATATGCCA TAAAATGACA TTCATATTT AAAGAGTTTT 120
 TAAAAACTCT TGTATTACCA TGCCATAATT TGAAACCTTA TTTCACTGAA TGAGAATGGT 180
 30 ATCTGTTGTC CTCATTTTT CATTTTATC CTTAACATT TCCACCACAG CCAGTGCATA 240
 TAATGGCAAT GACACCCAGG GATGGAATGA TAAGTTCCAT CRCMGCTCAG TCAAGACGCA 300
 GACTTGATGT GGCCCCAACCA ACAGTCAATA ATGGAGTCTC CAAAATAAG CTCTATAGGA 360
 35 AAGGTAAATA CCCGCTGCAC AAGAAACAC AGCACTAGG TTCTAACCCC ATCTCTATGA 420
 AGAGCTTGCT GGGAGAGTT TGACATTWAA CAATCTGTCT GATKGCCAAAT TTYTTCTTC 480
 40 TATAAAATGA TAATGTTKGA YTCAAAGATC CAAAGTCAAT TCATGGTCTA AAACTTAATG 540
 ATTTTTTTAG GTTTTGKGAC ATTICACTGT ACAGTGTAGT AATTTATATC TTATTTCCC 600
 ACTAATTTAG AAAATATYT AAATGATCCT TAATTGGCAA TGGGTCTAA GAATTTGTGTT 660
 45 TTAAATCCCT GTTACCCAAA AGAGCCCTTT TTGTATCTC GCAGTAGTTA CAAGGATCTT 720
 TCTAAATCTT AAAAAAAA AAAAAAGAAA GAAAGAAAAG AAAAGAAAAA AAGTCAGCCG 780
 50 GGCGTGGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAC CAAGGTGGAC AGATCACCG 840
 GTCAGGAGAT GGAGACCATC CGGGCCAACA TGGAGAAACC CTGTCCTAC TAAAAAAA 900
 55 AAAAACTCGA GGGGGGCCCCG GTACCCAATN CGCCGGCTAG TGGTCGTAAA ACAATCAA 959

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1446 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

10	CGGGGCAGGG CTGTGTGGCA CGGCCAGGGA GCGGGCCCAC CTGAGTCACT TTATTGGGTT	60
15	CACTAACAC TTTCTTGCTC CCTGTTTCT CTTCTGTGGG ATGATCTCAG ATCCAGGGGC	120
20	TGGTTTGGG GTTTCCITGC TTGTGCCAAG GGCTGGACAC TGCTGGGGGG CTGGAAAGCC	180
25	CCTCCCTTCC TGTCCCTCTG TGGCCTCCAT CCCCTCATGG GTGCTGCCAT CCTTCCTGGA	240
30	GAGAGGGAGG TGAAAGCTGG TGTGAGCCCA GTGGGTTCCC GCCCACTCAC CCAGGAGCTG	300
35	GCTGGGCCAG GACCGGGAGA GGGAGCACTG CTGCCCTCCT GGCCCTGCTC CTTCCGCAGT	360
40	TAGGGGTGGA CCGAGCCTCG CTTTCCCCAC TGTTCTGGAG GGAAGGGGAA GGAGGGGGTC	420
45	TTCAGGCTGG AGCCAGGCTG GGGGTGCTGG GTGGAGAGAT GAGATTTAGG GGGTGCCTCA	480
50	TGGGTGGGC AGGCCTGGGG TGAAATRAGA AAGGCCCAGA ACGTGCAGGT CTGCGGAGGG	540
55	GAAGTGTCTT GAGTGAAGGA GGGGACCCCCC ATCCTGGGGG ATGCTGGGAG TGAGTGAAGTG	600
60	AGATGGCTGA GTGAGGGTTA TGGGGACCT GAGGTTTAT GGGCTGTGT ATCCCCCTTCT	660
65	CCCGGCCCCA GCCTGCCTCC CTCCCTGCCG CCTGGCCCAC AGGTCTCCCT CTGGTCCCTG	720
70	TCCCTCTGGT GGTTGGGAT GGACGGCAG CAAGGGGTGT AATGGGGCTG GGTTCTGTCT	780
75	TCTACAGGCC ACCCCGAGGT CCTCAGTGGT TGCCCTGGGA GCCGGACGGG CCTCCTGAGG	840
80	GGTACAGGTT GGGTGGGCC TCCCTGAGGG TCTGGGTCA GGCTTTGGCT CTGCTGCCTC	900
85	TCAGTCACCA AGTCACCTCC CTCTGAAAAT CCAGTCCCTT CTTTGGATGT CCTTGTGAGT	960
90	CACTCTGGGC CTGGCTGTCTG TCCCTCCTCA CCTTCCTGTG CCTGGGACAA GGGTCAAGCC	1020
95	AGGATGGGCC CAGGCCTGGG ATCCCCCACC CCAGGACCCC CAGGCCCCCT CCCCTGCTGC	1080
100	TTTGGGGGG GCAGGGCAGA AATGGACTCC TTTTGGGTCC CCGAGGTGGG GTCCCCCTCCC	1140
105	AGCCCTGCAT CCTCCGTGCC STAGACCTGC TCCCCAGAGG AGGGGCCTTG ACCCACAGGA	1200
110	CGTGTGGTGG CGCCTGGCAC TCAGGGACCC CCAGCTGCC CAGCCCTGGT CTCTGGCGCA	1260
115	TCTCTTCCCT CTTGTCCGA AGATCTGCGC CTCTAGTGCC TTTTGAGGGG TTCCCATCAT	1320
120	CCCTCCCTGA TATTGTATTG AAAATATTAT GCACACTGTT CATGCTTCTA CTAATCAATA	1380
125	AACGCTTTAT TTAAAGCCAA AAAAAAAA AAAAAACTCG AGGGGGGCC CGTACCCAAT	1440
130	TOGCCA	1446

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1471 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

	CAAAAAATAA TAATGATAAT TTAAAATAAA TAAGTAACCA ATAAAAAGAT TTTATATCCC	60
15	AGCTTATGA TGTGGTTGG CAAGGCTAGA TAAAAAGATG TTAGAATGAA AGAACATATT	120
	TTTAGTGATA TGTAAATGAA GGATTCTACA ATAGTCATAT ATTTTATAT GAATGAATGT	180
	TGGGTTGGGC TGGAGAGGTA TGIGTGTGTA AATATAAAGG TCTCACATTG AGAGTATAGC	240
20	TCTGAAATAA TGGAACTCAT GTCTACAATT CAACATGCAT CTGTATAGTT ACATCTCATG	300
	TAAATATACA CAGACATATT TTGCAGCCAG TAATIGACAG TTAATGTCCA AAACAGGTGA	360
	TTGATAGGTA ACAGAAATTA GATAACCACC AATTTGCCA AAGAGAAAGA CTAGAAGGAC	420
25	TAAAAGCAGT TGAATGTATG GTACTGACAT TGTCTAAGC AGTCTGATAA CCAGTTTATT	480
	GAAACGTGTG CATTAAACAGA GAATTTAATT TTAAACCCAT AATTCTCCT ATCCATTAAA	540
30	ATATTATAAT TGTTAGTAGT ATGAAACCAA CAGGAAATGT TTTTAATCA TTAGTGAGG	600
	TGATTCAATTG GTTTCATGGG CAAACACTAT CCAGGAAAAG CCTTGCTTGC CTGTTTCCC	660
	AAGAGCTCTA AGAAATAGAA TCAAGTGTAA AATGGTTCA ACCATTCAAG ATTCTCTGTC	720
35	ACTCTCTCA ACCCGATCT TCCCTGTTACT ACTGATGTTT GAAACCTGT CATTAGCCCC	780
	GGCCTGGTTA AAGCCCCCTCA GAGTCACCTC TCATTCTAG CAATAGAATT CAACCCCAAG	840
40	TGGTTGATGG TGTCCCCAGC ACAGCCGAGA GACCTGATCT CTGGATTCA TGCTTTAGC	900
	TCTTCGAGTT TACCTTAAGA TACCTCGGG CAATATTTT AACCAACCCA AAAGCTCTTC	960
	AGGTCAATTTC TGAAGAGGAC AAGGTGAATC TTGGCTTGGG ACACCATTTC TGGCTCTTG	1020
45	CTACTGAATG AATCAGAAAG GAATTTTTC TGAAGAGCAT TAGAAAGTAA AGGAGATGTT	1080
	AAAATAAGTT CTTGAAGTAT GTTTATATT TATCTAAAC ACTGATTTA AAAGTTTACA	1140
50	TTCAAATGTG TATTCAAAG AAGTACTGAT TTGTAATTAT TATAGTTTGT GTGTATCATC	1200
	CCCTTTAAC CGTGCCTAAC AACTGTACTT AAATTTGTT TTCCTAGTGT AACAAATGTT	1260
	TCCCATAAGA TTTCTAGAG CCAAATAATG GGAGTGAAA ATTCTTAAG TGTTATATAA	1320
55	GAAAATATAT TAGAAAATCA GCTTGGATT ATACGATTTC TAAAATATAC TAATACAGAA	1380
	TCCTCAGTAA TATGTTTGA ATTGGATTTT TTCTCAGAAC TGTTACATAA TAAATAATAC	1440
60	ATCAACCAGA AAAAAAAA AAAAAATTN C	1471

5 (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1402 base pairs
 (B) TYPE: nucleic acid
 10 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

15	AGGGACGTCT TGCCTGAGGA GATGCCATT TCTGTCTGG RTTACCTCA CTGGCGTGGTG	60
	CATGAGCTGC CAGAGCTGAC GCGGGAGAGT TTGGAAGCAG GTGACAGTAA CCAATTTCGC	120
20	TGGAGGAACC TCTTTCTTG TATCAATCTG CTTCCGATCT TGAACAAGCT GACAAAGTGG	180
	AAGCAATTCAA GGACAATGAT GCTGGTGGTG TTCAAGTCAG CCCCCATCTT GAAGCGGGCC	240
	CTAAAGGTGA AACAAAGCCAT GATGCAGCTC TATGTGCTGA AGCTGCTCAA GGTACAGACC	300
25	AAATACTTGG GCGGGCAGTG GCGAAAGAGC AACATGAAGA CCATGCTGC CATCTACCAAG	360
	AAGGTGCGGC ATCGGCTGAA CGACGACTGG GCATACGGCA ATGATCTTGA TGCCCGGCCT	420
	TGGGACTTCC AGGCAGAGGA GTGTGCCCTT CGTGCCAAC A TTGAACGCCCTT CAACGGCCGG	480
30	CGCTATGACC GGGCCCACAG CAACCCCTGAC TTCTGOCAG TGGACAAC TG CCTGCAGAGT	540
	GTCCTGGGCC AACGGGTGGA CCTCCCTGAG GACTTCAGA TGAACATATGA CCTCTGGTTA	600
35	GAAAGGGAGG TCTTCTCCAA GCCCATTTC TGGGAAGAGC TGCTGCAGTG AGGCTGTTGG	660
	TTAGGGGACT GAAATGGAGA GAAAAGATGA TCTGAAGGTA CCTGTGGGAC TGTCTTAGTT	720
	CATTGCTGCA GTGCTCCCAT CCCCCACCAAG GTGGCAGCAC AGCCCCACTG TGTCTTCCGC	780
40	AGTCTGTCCT GGGCTTGGGT GAGCCAGCT TGACCTCCCC TTGGTCCCA GGGTCTGCT	840
	CCGAAGCAGT CATCTCTGCC TGAGATCCAT TCTTCCCTTA MTCCCCCAM CCTCCCTCT	900
45	TGGATATGGT TGGTTTGGC TCATTTACA ATCAGCCAA GGYTGGAAA GCTGGAATGG	960
	GATGGGAACC CCTCCGGCGT GCATCTRAAT TTCAGGGTC ATGCTGATGC CTCTCGAGAC	1020
	ATACAAATCC TTGCCTTGT CAGCTTGCAA AGGAGGAGAG TTTAGGATTA GGGCCAGGGC	1080
50	CAGAAAGTCG GTATCTTGGT TGTGCTCTGG CGTGGGGGTG GGGTGTCT GATGTTATTC	1140
	CAGCCCTCTG CTACATTATA TCCAGAAAGTA ATTGGGGAGG CCTCCCTCAGC TGCCTCAGCA	1200
55	CMTTGATTTT GGACAGGGAC AAGGTAGGAA GAGAAGCTTC CCTTAACCAG AGGGGCCATT	1260
	TTTCCTTTG GCTTTCGAGG GCCTGTAAT ATCTATATAT AATTCCTGIGT GTATTCCTG	1320
60	TCATGTTGGG GTTTTAATG TGATTGTGTA TTCTGTTAC ATTAAAAAGA AGCAAAAATA	1380

ATAAAAAAA AAAAAAAA CT

1402

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1047 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

15	GGCACAGGGG ACTACAGGCA CCCACGACCA TACCCAGCTA ATTTTGTAT TTTTTGTAG	60
20	AGATGGGGTT TCACGATGTC GCCCAGGCTG GTCTTGAACt CCTGGGCTTG AGCGATCTC	120
25	CCATCTTTC CCATCTGGCCT CCTAAAGTGC TGGGACTGCCA GGCAATGAGCC ACCATGCCA	180
30	GCCAAGATTG TTATTGATTA CCATGTTGCT TCAAGAAGCC AAGCCAGTTT CCAATATTCC	240
35	CCATTTGCTG GAGTCCTGGT ACTTTGGTA GAAGCAACTG GTAAATTGTT AATTGGAACA	300
40	TTGGTGGTG TAGATAACCA CGTATGGCCA AACCTAGAGC ATCTAGGCTC ACAATTACTA	360
45	TCTGACTTG ATAACAAGTG TTCTGATATT AACCTGAAAA TGGGAATAAT GCCAAATCTG	420
50	TGTAACCTAA CATCTATATA CACAGTGGGG AGAACTGAAG TTATTAACC TGGAATCTCT	480
55	GTGATCAAGG CTAACAGTAG TTATCTAAGA AGCAAAGGAC CTACAATTCT TAGACTTGG	540
60	GTCATATTCT TTAAGGACGT GTTCTGAAAC TATATCAAGC ATCTGGTTTC CACGTATTTC	600
65	TCCCTCAGAA ATTATGAAGT ACAAGTAAGA ATGAAGGTAC AGGGTAAGAC ACATGCTGCT	660
70	TTCTTGCTCT TGAGTGGAGA CAGTTTCCA GCCATCTTAA CCCCTIWACA CAAAACAATT	720
75	TGTTTTAT AGCAAATAAG TGACTCAACA TAATTCAAT ATGATGTTA TCCACCAGTA	780
80	CTTTCCTTTC AGCTTCTAGT CCCATAARTG GTTTGTGAAG TCATCGTTA CATTAGCCAA	840
85	GATAGGCCTA GACTTGAAGT CTAGAACGTT TTTCCCACCA TATGCCAAAG TAGAATGTGG	900
90	GTATCTCAGG GTCAATTG TTGTTCAATT TCCCACCTGT ACAGTTGTTA TGATTCACTT	960
95	TCCATTATGTG TCTAATAAT CTTGTTCCAT GAAATGATCA AAAAAAAA AAAAAAAACT	1020
100	CGAGGGGGGG CCCGGTACCC AAATCGC	1047

55 (2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5	TTGGAAAGGG TCTAGCTCTT TCTCATTCAAC CAACTATATT AGAACCAC TT GAGGGAAATT	60
	TACCACTCCA AATCCAAAGC AATGAACAGT CTTTCTGGA TGATTTTATT GCCTGTGTCC	120
10	CAGGATCAAG TGGTGGAAAGG CTTGCAAGGT GGCTTCAGCC AGATTCAATAT GCGGATCCCTC	180
	AGAAAACATC TTTGATCCTG GAATAAGGAT GATATTCTTT GTGGTGGGCC TACCACCATA	240
15	ACTGTTCAAA CAAAAGACCA GTATGGGAT GTGGTACATG TTCCCAATAT GAAGGTAATT	300
	ATAACTGGAT TAAATTAGCA GACATCTATA TACTGGCTGC AATGACTGAT AAAATTTAG	360
20	AAATGCCAAG TGCTGAGRGT CCATTGTTTC TACCTCTTT ATATAAAGGG TGATGCTGAA	420
	AGTTTGTAA AATGACTTGT TTATATTAAT TAGTCCCCAA GTGTCCAAGT TACACCTGTT	480
25	TTTTTGTTGA GTTTGGTCTT TACATTTTGC TACCTGTTAC GGGGACTCAA AGGAGGGATA	540
	AGAAAGTATC CATCTAAAGA GTGCTAGACA CATACTGTA AGCCCCCTCAA TATGTATTGA	600
30	TTGAATAAT GCATGAAAGA ATACATTTTT AAATTTTGTG TATAGTTTG AAAGACTCAA	660
	GTACGTTCTG TGTTTGGTAT TACTGAAACC ACATTTTAAA AATAACACTC ATTAAGTTAG	720
35	AAATATATGA GTTTAGATTG TAAAAGAATG AGGAATTGAA ATAGTTGTAT ACCATATTGA	780
	TGAATATAGA GTTTTAGGA TACCTCTTAC CTGAAATATT AATAATAATG TTTNCAGAGC	840
	ATATTATACA TAATTATTTG TGATTTAAC TGTAAATATG AATATCTCAT TTAAAACTTT	900
40	TATTCTGAA AAAATTATAT TGAATAAAAT TTTATATAGG CAGTCCCCAG CCCTTTCCCTC	960
	CTTCAAAGTT GTCTTATAGA GTGATTGGTT	990

40

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1208 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

	TAATCGCTAC TATAGGGAAA GCTGGTCGCT GCAGGTACOG GTCCCGAATT CCGGGTCGAC	60
55	CCACGCGTCC GAGCGAAATG GCGCCTCCGG CCCCGGGCCC GGCGCTCCGGC CGCTCCGGGG	120
	AGGTAGACGA GCTGTTGAC GTAAAGAACG CCTTCTACAT CGGCAGCTAC CAGCAGTGCA	180
	TAAACGAGGC GCASGGGTGA AGCTRTCAAG CCCAGAGAGA GACGTGGAGA GGGACGTCTT	240
60	CCTGTATAGA CGGTACCTGG CGCAGAGGAA GTTCGGTGTG GTCCCTGGATG AGATCAAGCC	300

	CTCCTCGGCC CCTGAGCTCC AGGCCGTGCG CAATGTTGCT GACTACCTCG CCCACGAGAG	360
	TCGGAGGGAC AGCATCGTGG CCGAGCTGGA CCGAGAGATG AGCAGGAGCK TGGACGTGAC	420
5	CAACACCACC TTCCCTGCTCA TGGCCGCTC CATCTATCTC CACGACCAGA ACCCGGATGC	480
	CGCCCTGCGT GCGCTGCACC AGGGGGACAG CCTGGAGTGC ACAGCCATGA CAGTGCAGAT	540
10	CCTGCTGAAG CTGGACCGCC TGGACCTCGC CGCGAAGGAG CTGAAGAGAA TGCAGGACCT	600
	GGACGAGGAT GCCACCCCTCA CCCAGCTCGC CACTGCCCTGG GTCAGCCTGG CCACGGGTGG	660
	TGAGAACGCTG CAGGATGCCT ACTACATCTT CCAGGAGATG GCTGACAAGT GCTCGCCAC	720
15	CCTGCTGCTG CTCATGGGC AGGCGGCCTG CCACATGGCC CAGGGCCGCT GGGAGGGCCG	780
	TGAGGGCCTG CTGCAGGAGG CGCTAGACAA GGATAGTGGC TACCCRGAGA CGCTGGTCAA	840
20	CCTCATCGTC CTGTCCCAGC ACCTKGGCAA GCCCCCTGAG GTGACAAACC GATACTGTC	900
	CCAGCTGAAG GATGCCACA GGTCCCATCC CTTCATCAAG GAGTACCAAG CCAAGGAGAA	960
	CGACTTTGAC AGGCTGGTGC TACAGTACGC TCCCAGCGCT GAGGCTGGCC CAGAGCTGTC	1020
25	AGGACCATGA AGCCAGGACA GAGGCCAGGA GCCAGCCCTG CAGCCCTCCC CACCCGGCAT	1080
	CCACCTGCAT CCCCTGGGG CAGGAGCCCA CCCCCAGCAC CCCCATCTGT TAATAAATAT	1140
30	CTCAACTCCA RGGTGTTCGA CCTGAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1200
	AAAAAAAAA	1208

35

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 1922 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

	GTGCTGCGCT ACTGAGGAGC GCCATGGAGG ACTCTGAAGC ACTGGCTTC GAACACATGG	60
	GCCTCGATCC CCGGCTCTT CAGGCTGTCA CCGATCTGGG CTGGTCGCGA CCTACGCTGA	120
50	TCCAGGAGAA GGCCATCCA CTGGCCCTAG AAGGGAAGGA CCTCCTGGCT CGGGCCCGCA	180
	CGGGCTCCGG GAAGACGGCC GCTTATGCTA TTCCGATGCT GCAGCTGGTG CTCCATAGGA	240
55	AGGCGACAGG TCCGGTGGTA GAACAGGCAG TGAGAGGCCT TGTTCTTGT CCTACCAAGG	300
	AGCTGGCACAG GCAAGCACAG TCCATGATTC AGCAGCTGGC TACCTACTGT GCTCGGGATG	360
60	TCCGAGTGGC CAATGTCTCA GCTGCTGAAG ACTCAGTCTC TCAGAGAGCT GTGCTGATGG	420

	AGAAGCCAGA TGTGGTAGTA GGGACCCAT CTCGCATATT AAGCCACTTG CAGCAAGACA	480
	GCCIGAAACT TCGTGACTCC CTGGAGCTTT TGGTGGTGGG CGAAGCTGAC CTTCTTTTTT	540
5	CCTTGGCTT TGAAGAAGAG CTCAAGAGTC TCCCTCTGTCA CTGCCCCGG ATTTACCAGG	600
	CTTTCTCAT GTCAGCTACT TTAAACGAGG ACGTACAAGC ACTCAAGGAG CTGATATTAC	660
10	ATAACCCGGT TACCCCTTAAG TTACAGGAGT CCCAGCTGCC TGGGCCAGAC CAGTTACAGC	720
	AGTTCAAGT GGTCTGTGAG ACTGAGGAAG ACAAAATTCCT CCTGCTGTAT GCCCTGCTCA	780
	AGCTGTCAATT GATTGGGGC AAGTCTCTGC TCTTTGTCAA CACTCTAGAA CGGAGTTACC	840
15	GGCTACGCCT GTTCTTGGAA CAGTCAGCA TCCCCACCTG TGTGCTCAAT GGAGAGCTTC	900
	CACTGCGCTC CAGGTGCCAC ATCATCTCAC AGTCAACCA AGGCTTCTAC GACTGTGTCA	960
20	TAGCAACTGA TGCTGAAGTC CTGGGGCCC CAGTCAGGG CAAGCGTCGG GGCGGAGGGC	1020
	CNAAAGGGGA CAAGGCCTCT GATCCGAAG CAGGTGTGGC CCGGGGCATA GACTTCCACC	1080
	ATGIGTCTGC TGTGCTCAAC TTGATCTTC CCCCCACCCC TGAGGCCTAC ATCCATCGAG	1140
25	CTGGCAGGAC AGCACCGCCT AACAAACCCAG GCATAGTCCT AACCTTTGTG CTTCCCACGG	1200
	ACCAAGTCCA CTTAGGCAAG ATTGAGGAGC TTCTCAGTGG AGAGAACAGG GGCCCCATTG	1260
	TGCTCCCTA CCAGTTCCGG ATGGAGGAGA TCGAGGGCTT CGCGTATCGC TGCAGGGATG	1320
30	CCATGCGCTC AGTGAUTAAC CAGGCCATTG GGGAGGCAAG ATTGAAGGAG ATCAAGGAAG	1380
	AGCTTCTGCA TTCTGAGAAG CTTAAGACAT ACTTTGAAGA CAACCCTAGG GACCTCCAGC	1440
35	TGCTGGGCA TGACCTACCT TTGCACCCCG CAGTGGTGA GCCCCACCTG GGCCATGTTC	1500
	CTGACTACCT GGTTCTCCT GCTCTCCGTG GCCTGGTRCG CCCTCACAAAG AAGCGGAAGA	1560
	AGCTGTCTTC CTCTTGTAGG AAGGCCAAGA GACCAAAGTC CCAGAACCCCA CTGCGCAGCT	1620
40	TCAAGCACAA AGGAAAGAAA TTCAGACCCA CAGCCAAGCC CTCCCTGAGGT TGTGTTGGCCT	1680
	CTCTGGAGCT GAGCACATTG TGGAGCACAG GCTTACACCC TTCTGGACA GGCGAGGCTC	1740
45	TGGTGCTTAC TGCACAGCCT GAACAGACAG TTCTGGGCC GGCAGTGCTG GGCCCTTAC	1800
	CTCCCTGGCA CTTCCAAGCT GGCACTCTGC CCCTTGACAA CAGAATAAAA ATTTAGCTG	1860
	CCCCAAAAAA AAAAAAAA AAAAAAAACTC GAGGGGGGGC CCGTACCCAA TTCGCCCTAT	1920
50	AA	1922

55

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1951 base pairs
- (B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5	TCGTCCCCAG ACGGGGCTGA GCCCCAGGCG SAGGGTGGCG GGGGAGCCTG GGGGAGCCGC	60
10	CGCCACCTCC ACGGGCCTCT CTGAGGCTCG ACACCAGCGC CCTGTCCTAT GACTCTGTCA	120
15	AGTACACGCT GGTGGTAGAT GACCATGCAC AGCTGGAGCT GGTGAGCCTG CGCCGTGCTT	180
20	CGGAGACTAC AGTGACGAGA GTGACTCTGC CACCGTCTAT GACAACGTG CCTCCGTCTC	240
25	CTCGCCCTAT GAGTCGGCCA TCGGAGAGGA ATATGAGGAG GCCCCGCGGC CCCAGCCCCC	300
30	TGCCCTGCCTC TCCGAGGAAC TCCACGCCTG ATGAACCCGA CGTCCATTTC TCCAAGAAAT	360
35	TCCCTGAACGT YTTCATGAGT GGCGCTCCC GCTCCTCCAG TGCTGAGTCC TTGGGGCTGT	420
40	TCTCCTGCAT CATCAACGGG GAGGAGCAGG AGCAGACCA CCGGGCCATA TTCAGGTTTG	480
45	TGCCTCGACA CGAACGCAA CTTGAGCTGG AAGTGGATGA CCCCTCTGCTA GTGGAGCTCC	540
50	AGGCTGAAGA CTACTGGTAC GAGGCCTACA ACATGCGCAC TGGTGCCCCG GGTGTCTTTC	600
55	CTGCCTATTA CGCCATCGAG GTCACCAAGG AGCCCGAGCA CATGGCAGCC CTGGCCAAAA	660
60	ACAGTGACTG GGTGGACCAG TTCCGGGTGA AGTTCTGGG CTCAGTCCAG GTTCCCTATC	720
65	ACAAGGGCAA TGACGTCTC TGTGCTGCTA TGCAAAAGAT TGCCACCACC CGCCGGCTCA	780
70	CCGTGCACTT TAACCCGCC TCCAGCTGIG TCCTGGAGAT CAGCGTGGGG GGTGTGAAGA	840
75	TAGGCCTCAA GGCGATGAC TCCCAGGAGG CCAAGGGAA TAAATGTAGC CACTTTTTC	900
80	AGTTAAAAAA CATCTCTTTC TCGGGATATC ATCCAAAGAA CAACAAGTAC TTTGGGTTCA	960
85	TCACCAAGCA CCCCAGCGAC CACCGGTTTG CCTGCCACGT CTTTGTGTCT GAAGACTCCA	1020
90	CCAAAGCCCT GGCAAGAGTCC GTGGGGAGAG CATTCCAGCA GTTCTACAAG CAGTTTGTGG	1080
95	AGTACACCTG CCCCACAGAA GATATCTACC TGGAGTAGCT GTGCAGCCCC GCCCTCTGCG	1140
100	TCCCCCAGCC CTCAGGCCAG TGCCAGGACA GCTGGCTGCT GACAGGATGT GGCACTGCTT	1200
105	GAGGAGGGGC ACCTGCCACC GCCAGAGGAC AAGGAAGTGG GGCGCTGGCC CAGGGTAGGG	1260
110	GAGGGTGGGG CAATGGGGAG AGGCAAATGC AGTTTATTGT AATATATGGG ATTAGATTCA	1320
115	TCTATGGAGG GCAGAGTGGG CTGCTGGGG ATTGGGAGGG ACAGGGCTTG GGGAGCAGGT	1380
120	CTCTGGCAGA GAAGGATGTC CGTTCAGGA GCACACGGCC CTGCCCATC CTGGGCCCTA	1440
125	CCTCCCCCTGC CAGGGCTCGG GCGCTGTGGC TCCCTGCCTTG ATGAAGCCCG TGTCCCTGCCT	1500
130	TGATGAAGCC TGTGCCACCT GCAAGTGCC CCCCCGCCCC TGCCCCAACC CCCACCGAAG	1560
135	AGCCCTGAGC TCAGGCTGAG CCCAGCCACC TCCCAAGGAC TTTCCAGTGA GGAAATGGCA	1620
140	ACACGTGGAG GTGAAGTCCC TGTCTCAGC TCCGTCTATCT GCGGGGCTTC TGGGTGGCTC	1680

CTGCCACTGA CCTCACCGGC ATGCTGGCCT GTGGCAGGCC TAGGACCTCA GGCAGGGAGG	1740
AGGAGCTGCC GCAAGGCCT GTCCCACAG AAGAGGGAGG CTTCCCTGACT GACACAGGCC	1800
5 AGCCCCATCT TGGTCCTGTC ACCCTGGCC CAACTATTAA AGTGCCTTT CCTGTCAAAA	1860
AAAAAAAAA AAAATCGGGG GGGGCCGGA ANCCAATTTC CCCCAAAAAG GCGGGTTATA	1920
10 AAAATTCCCN GGCGNTGT TTAAAAATTG G	1951

15 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3989 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25 GGCACAGGCC GCAGGGNACC TATGGGCGCA TATAGGTGT AATGAAACTG TAGTCTCAGT	60
TGGAAGCCTA GACATGAAAT GGGTCAGTGA GCAAGGCTCT ATTCCCTAGTC TCCAGCCATG	120
30 CCTGTGGAAC CTGARCCRC TCTCAGCACA TTGGACCCAG GCAGATGYAA AAAATTCAAA	180
GAACTATGAT TTGGACTCAA GGGTTTGTAG ATTTCCCTCT TCATTTCTAAT TTCAGTGTCT	240
AAAATTCTTG CATCCRTGAA CGAGCTGGC ATTTGATGAG ACAGGGCYGA ATACTGCAGT	300
35 TTTCCTCCTA GAAATCATCT GGGCACTTTT CTTTGAACCTG ATGGGAACAA TGGGCATAAA	360
CTGTTTGCAC AAACTTGGGA TAARTGATTT TGGGATAACG ATCTACCAGA ATGGGGATAT	420
40 TTCACCCCTG GTTCTGAGAT GCAAACCAAA GAATATCATG ACCAGCTTTC AGGCCTCCTG	480
AAGTATATCT CTCACATTGT CCTGTTCTCA TGCTGAGGAG CCTGAGATCC CTGTTGTTGG	540
ATTAGACAGT GGACTGTAT GGGTGTAGGT GAATTGGCTT ATTTTGTCTG TCCCCTGTCTG	600
45 AATGTATTGC AGGAAYTAAA AAGGACCAAG AAGAGGAAGA AGACCAAGGC CCACCATGCC	660
CCAGGCTCAG CAGGGAGCTG CTGGAGGTAG TAGAGCCTGA AGTCTTGCAG GACTCACTGG	720
ATAGATGTTA TTCAACTCCT TCCAGTTGTC TTGAACAGCC TGACTCCTGC CAGCCCTATG	780
50 GAAGTTCCCTT TTATGCATTG GAGGAAAAC ATGTTGGCTT TTCTCTTGAC GTGGGAGAAA	840
TTGAAAAGAA GGGGAAGGGG AAGAAAAGAA CGGGAAAGAAG ATCAAAGAAG GAAAGAAGAA	900
55 GGGGAAGAAA AGAAGGGAA GAAGATCAA ACCCACCATG CCCCAGGCTC AGCAGGGAGC	960
TGCTGGATGA GAAAGRGCCT GAAGTCTTGC AGGACTCACT GGATAGATGT TATTCAACTC	1020
60 CTTCAAGTTGT GTTGAACGTGT GTGACTCATG CCAGCCCTAC AGAAGTGCCT TTTATGTATT	1080

	GGAGCAACAG CATGTTGGCT TGGCTGTTGA CATGGATGAA ATTGAAAAGT ACCAAGAAGT	1140
	GGAAGAAGAC CAAGACCCAT CATGCCCGAG GCTCAGCAGG GAGCTGCTGG ATGAGAAAGA	1200
5	GCCTGAAGTC TTGCAGGACT CACTGGATAG ATGTTATTG ACTCCTTCAG GTTATCTTGA	1260
	ACTGCCTGAC TTAGGCCAGC CCTACAGCAG TCCKGTITAC TCATTGGAGG AMCAKTACCT	1320
10	TGGCTTKKCT CTTGACGTGG ASAAATTGAA AAGAAGGGGA AGGGGAARAA AAGAAGGGGA	1380
	AGAACATCAA AGAAGGAAAG AAGAAGGGGA AGAAAAGAAG GGGAAACAAGA TCAAAACCCA	1440
	CCATGCCCA GGCTCAGCAG GGAGCTGCTG GATGAGAAAG GGCTGAAGT CTTGCAGGAC	1500
15	TCACTGGATA GATGTTATTIC AACTCCCTCA GGTTGTCCTTG AACTGACTGA CTCATGCCAG	1560
	CCCTACAGAA GTGCCCTTTA YRTATTGGAG CAACAGYGTG TTGGCTTGGC TGTTGACATG	1620
	GATGAAATTG AAAAGTACCA AGAAGTGGAA GAAGACCAAG ACCCATCATG CCCCAGGCTC	1680
20	AGCAGGGAGC TGCTGGATGA GAAAGAGCCT GAAGTCTTGC AGGACTCACT GGATAGATGT	1740
	TATTCGACTC CTTCAGGTTA TCTTGAACCTG CCTGACTTAG GCCAGCCCTA CAGCAGTGCT	1800
25	GTTTACTCAT TGGAGGAACA GTACCTTGGC TTGGCTCTTG ACGTGGACAG AATTAAGGAG	1860
	GACCAAGAAG AGGAAGAAGA CCAAGGCCA CCATGCCCA GGCTCAGCAG GGAGCTGCTG	1920
	GAGGTAGTAG AGCCTGAAGT CTTGCAGGAC TCACTGGATA GATGTTATTIC AACTCCCTCC	1980
30	AGTTGTCCTTG AACAGCCTGA CTCCCTGCCAG CCCTATGGAA GTTCCCTTTA TGCAATTGGAG	2040
	GAAAAACATG TTGGCTTTTC TCTTGACGTG GGAGAAATTG AAAAGAAGGG GAAGGGGAAG	2100
35	AAAAGAAGGG GAAGAAGATC AAMGAAGRAA AGAAGAAGGG GAAGAAAAGA AGGGGAAGAA	2160
	GATCAAAACC CACCATGCC CAGGCTCAAC GGCGTGTGA TGGAAAGTGGAA AGAGCSTGAA	2220
	GTCTTACAGG ACTCACTGGA TAGATGTTAT TCGACTCCGT CAATGTACTT TGAACACTACCT	2280
40	GACTCATTCC AGCACTACAG AAGTGTGTTT TACTCATTTG AGGAACAGCA CATCAGCTTC	2340
	GCCCTTTACG TGGACAATAG GTTTTTACT TTGACGGTGA CAAGTCTCCA CCTGGTGTTC	2400
45	CAGATGGGAG TCATATTCCC ACAATAAGCA GCCCTTASTA AKCCGAGAGA TGTCAATTCC	2460
	GCAGGCAGGA CCTATAGGCA MGTGAAGATT TGAATGAAAG TACAGTTCCA TTTGGAAGCC	2520
	CAGACATAGG ATGGGTCACTT GCCCATGGCT CTATTCCAT TCTCAAACCA TGCCAGTGGC	2580
50	AACCTGTGCT CAGTCTGAAG ACAATGGACC CACGTTAGGT GTGACACGTT CACATAACTG	2640
	TGCAGCACAT GCCGGGAGTG ATCAGTCRGA CATTTTAATT TGAACCACGT ATCTCTGGT	2700
55	AGCTACAAAA TTCCCTCAGGG ATTTCATTTC GCAGGCATGT CTCTGAGCTT CTATACCTGC	2760
	TCAAGGTCAK TGTCACTTT GTGTTTAGCT CATCCAAAGG TGTACCCCTG GTTCAATGA	2820
60	ACCTAACCTC ATTCTTTGTG TCTTCAGTGT TGGCTTGTGTT TAGCTGATCC ATCTGTAACA	2880

	CAGGAGGGAT CCTTGGCTGA GGATTGTATT TCAGAACAC CAACTGCTCT TGACAATTGT	2940
	TAACCCGCTA GRCTCCTTG GTTAGAGAAG CCACAGTCCT TCAGCCTCCA ATTGGTGTCA	3000
5	GTACTTAGGA AGACCACAGC TAGATGGACA AACAGCATTG GGAGGCCITA GCCCTGCTCC	3060
	TCTCRATTCC ATCCTGTAGA GAACAGGAGT CAGGAGCCGC TGGCAGGAGA CAGCATGTCA	3120
10	CCCAGGACTC TGCCGGTGCA GAATATGAAC AAYGCCATGT TCTTGCAGAA AACGCTTAGC	3180
	CTGAGTTTCA TAGGAGGTAA TCACCAAGACA ACTGCAGAAT GTRGARCACT GAGCAGGACA	3240
	GCTGACCTGT CTCCCTCAC A TAGTCCATRT CACCACAAAT CACACAACAA AAAGGAGARG	3300
15	AGATAATTGT GGITCAAAAA AAGTAAAAAG ATAATGTAGC TGCATTTCCTT TAGTTATTTT	3360
	GARCCCCAAA TATTTCCCTCA TCTTTTGTGTT GTTGTCAKG ATGGTGGTGA CATGGACTTG	3420
20	TTTATAGAGG ACAGGTCAGC TGCTGGCTC AGTGATCTAC ATTCTGAAGT TGTCTGAAA	3480
	TGTCCTCATG ATTAATTCA GCCTAACGT TTTGCCGGGA ACACTGCAGA GACAATGCTG	3540
	TGAGTTTCCA ACCTYAGCCC ATCTGCGGGC AGAGAAGGTC TAGTTTGTCC ATCASCATTA	3600
25	TCATGATATC AGGACTGGTT ACTTGGTTAA GGAGGGTCT AGGAGATCTG TCCCTTTAG	3660
	AGACACCTTA CTTATAATGA AGTATTGGG AGGGTGGTTT TCAAAATTAG AAATGTCCTG	3720
30	TATTCCRATG ATCATCCGT AAACATTITA TCATTTATTA ATCATCCCTG CCTGTGTCTA	3780
	TTATTATATT CATATCTCTA CGCTGGAAAC TTTCTGCCTC AATGTTTACT GTGCCTTTGT	3840
	TTTGCTAGT GTGTGTGTGTT GAAAAAAAAA ACATTCTCTG CCTGAGTTTT AATTTTGTG	3900
35	CAAAGTTATT TTAATCTATA CAATTAAAAG CTTTGCCTA TCAAAAAAAA AAAAAAAAAA	3960
	AAAAAAAAAA AAAAGCGGA CGCGTGGGC	3989

40

(2) INFORMATION FOR SEQ ID NO: 29:

45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 3735 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:	
	CTGCTGTTCG CTGGCTGGGC TCCGCAGCAG GCTTGGCCAG CSGCTGACGG GTCGGGGGC	60
55	GGTTTGTTGTGTT GAACAGGCAC GCAGCTGCAG ATTTCATTCT GGTAGTGCAN CCCTCTCAAA	120
	GGTTGAAGGA ACTGATGTAA CAGGGATTGA AGAAGTAGTA ATTCCAAAAA AGAAAACCTTG	180
	GGATAAAAGTA GCCGTTCTTC AGGCACCTTGC ATCCACAGTA AACAGGGATA CCACAGCTGT	240
60	GCCTTATGTG TTTCAAGATG ATCCCTACCT TATGCCAGCA TCATCTTTGG AATCTCGTTC	300

	ATTTTACTG GCAAAGAAAT CCGGGGAGAA TGTGCCAAG TTTATTATTA ATTACATAACCC	360
5	CAAATATTTT CAGAAGGACA TAGCTGAACC TCATATACCG TGTTTAATGC CTGAGTACTT	420
	TGAACCTCAG ATCAAAGACA TAAGTGAAGC CGCCCTGAAG GAACGAATTG AGCTCAGAAA	480
	AGTCAAAGCC TCTGTGGACA TGTTTGATCA GCTTTGCAA GCAGGAACCA CTGTGTCTCT	540
10	TGAAACAAACA AATAGTCTCT TGGATTWIT GTGTTACTAT GGTGACCAGG AGCCCTAAC	600
	TGATTACCAT TTTCAACAAA CTGGACAGTC AGAACATTG GAAGAGGAAA ATGATGAGAC	660
15	ATCTAGGAGG AAAGCTGGTC ATCAGTTGG AGTTACATGG CGAGCAAAAA ACAACGCTGA	720
	GAGAATCTTT TCTCTAATGC CAGAGAAAAA TGAACATTCC TATTGCACAA TGATCCGAGG	780
	AATGGTGAAG CACCGAGCTT ATGAGCAGGC ATTAAACTTG TACACTGAGT TACTAAACAA	840
20	CAGACTCCAT GCTGATGTAT ACACATTAA TGCATTGATT GAAGCAACAG TATGTGCGAT	900
	AAATGAGAAA TTTGAGGAAA AATGGAGTAA AATACTGGAG CTGCTAAGAC ACATGGTTGC	960
	ACAGAAGGTG AAACCAAATC TTCAGACTTT TAATACCATT CTGAAATGTC TCCGAAGATT	1020
25	TCATGTGTTT GCAAGATCGC CAGCCTTACA GGTTTACGT GAAATGAAAG CCATTGGAAT	1080
	AGAACCCCTCG CTIGCAACAT ATCACCATAT TATTGCCCTG TTGATCAAC CTGGAGACCC	1140
30	TTAAAGAGA TCATCCTTCA TCATTTATGA TATAATGAAT GAATTAATGG GAAAGAGATT	1200
	TTCTCCAAAG GACCCGGATG ATGATAAGTT TTTCAGTCA GCCATGAGCA TATGCTCATC	1260
35	TCTCAGAGAT CTAGAACTTG CCTACCAAGT ACATGGCCTT TTAAAAACCG GAGACAACIG	1320
	GAAATTCAATT GGACCTGATC AACATCGTAA TTCTTATTAT TCCAAGTTCT TCGATTGAT	1380
	TIGTCTAATG GAACAAATTG ATGTTACCTT GAAGTGGTAT GAGGACCTGA TACCTTCAGC	1440
40	CTACTTTCCC CACTCCCCAA CAATGATACA TCTTCTCCAA GCATTGGATG TGGCAATCG	1500
	GCTAGAAAGTG ATTCCCTAAA TTGGAAAGA TAGTAAAGAA TATGGTCATA CTTTCCGCAG	1560
	TGACCTGAGA GAAGAGATCC TGATGTCAT GGCAAGGGAC AAGCACCCAC CAGAGCTTCA	1620
45	GGTGGCATTG GCTGACTGTG CTGCTGATAT CAAATCTGCG TATGAAAGCC AACCCATCAG	1680
	ACAGACTGCT CAGGATTGGC CAGCCACCTC TCTCAACTGT ATAGCTATCC TCTTTTAAG	1740
50	GGCTGGGAGA ACTCAGGAAG CCTGGAAAAT GTGGGGCTT TTCAAGGAACC ATAATAAGAT	1800
	TCCTAGAAAGT GAGTTGCTGA ATGAGCTTAT GGACAGTGCA AAAGTGTCTA ACAGCCCTTC	1860
55	CCAGGCCATT GAAGTAGTAG ACCTGGCAAG TGCCTTCAGC TTACCTATT GTGAGGGCCT	1920
	CACCCAGAGA GTAATGAGTG ATTTTCAAT CAACCAAGGAA CAAAAGGAAG CCCTAAGTAA	1980
	TCTAACTGCA TTGACCACTG ACAGTGATAC TGACAGCAGC AGTGACAGCG ACAGTGACAC	2040
60	CAGTGAAGGC AAATGAAAGT GGAGATTCAAG GAGCAGGAAT GGTCTCACCA TAGCTGCTGG	2100

	AATCACACCT GAGAACTGAG ATATAACCAAT ATTAAACATT GTTACAAAGA AGAAAAGATA	2160
5	CAGATTGGT GAATTTGTTA CTGTGAGGTA CAGTCAGTAC ACAGCTGACT TATGTAGATT	2220
	TAAGCTGCTA ATATGCTACT TAACCACATCA TTAATGCACC ATTAAAGGCT TAGCATTAA	2280
	GTAGCAACAT TGCGGTTTC AGACACATGG TGAGGTCCAT GCCTCTTGTC ATCAGGATAA	2340
10	GCCTGCACAC CTAGAGTGTC GGTGAGCTGA CCTCACGATG CTGTCCTCGT GCGATTGCC	2400
	TCTCCTGCTG CTGGACTTCT GCCTTTGTTG GCCTGATGTG CTGCTGTGAT GCTGGTCCTT	2460
15	CATCTTAGGT GTTCATGCAG TTCTAACACA GTTGGGGTTG GGTCAATAGT TTCCCAATT	2520
	CAGGATATT CGATGTCAGA AATAACGCAT CTTAGGAATG ACTAAACAAG ATAATGGCAG	2580
	TTTACGGCTGC ACAACTGGTA AAATGACTGT AGATAAAATGT TGTAATTAGT GTACACGTT	2640
20	GTATTTTGT TAATATAGCC GCTGCCATAG TTTTCTAATC TGAACAGCCA TGAATGTTTC	2700
	ATGTCCTCCCT TTTTTTTTG TCTATAGCTG TTACCTATT TAGTGGTTGA AATGAGAGCT	2760
	AGTGATGACA GAAGGATGTG GAATGTCTTC TTGACATCAT TGTGTATTGC TGGTAATCAA	2820
25	GTTGGTAACG ACTACTCTA GCAGCTCTTA CCACATGAC TTAAGTGGTC CTGGAAGGCA	2880
	GTAAGTGGAG GTTGGCAGCA TTCTGCCCT CATGAGGGCT TCTACCACTG ACCACTTTGC	2940
30	ACGTACCTGG CTCCCAGATT TACTTAGGTA CCCCACGAGT CGTCCACATA AGCAGCTTCA	3000
	TCTTTACCTT GCCAGAGTTG ACAATTATGG GATACTCTAG TCTACTTATA CTTGTGTTCC	3060
35	CATCTGTCTG CCATCCTCTG AAGGCCAGGA CCCAGTCATA CATCCTTAGA AACCAAAGTA	3120
	TGGTTTTGT TTTCTCTGG AATGTCAGGT CTTAAGGCAT TTAATTGAGG GACAAAAAAA	3180
	AAAAAAAGCC GATATAGTAG CTAGCTACTT AAGCATCCAT GGGTATTGCT CCATATCAA	3240
40	GCAGATTTGC AGGACAGAAA GAGTAAATTG GCCTTCAGTC TTGGTTTACA GCTTCCAAGG	3300
	AGAGCCCTGG CCACCTGAAA TGTAACTCG GTCCCTCCT GTCTCTAGTT CATCAGCACC	3360
45	TGCAGATGCC TGACTCTTGT TAGCCTTACT ATTCAATACA GTCCCTAGAT TCACGGTATG	3420
	CCTCTTCTTA TCCAGCCACC TATTCTGAAT CACCATGTTG CTCTGCAGCT AGAGTTGATA	3480
	GGAGAAAATC CATTGGGTA GATGGCCTAT GAATTGTTAG TAGACTTTCA AAATGAGTGA	3540
50	TTTGTAGCT TGGTACTTTT AAGTTGTTGG TACAGATCCT CCAAACCCAT ACTCTGAGCA	3600
	ATTAAC TGAAACATAG AGAAAATTAA GGCTCACAG GATGAGTC CATTCTCTG	3660
55	AAATGCTTAT TTTATCATAG TCTTTAGCCN CTACTATGAG TAAAATGTT CTTTCNGCCG	3720
	GGTGTGGTGA CTCAC	3735

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1667 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

10	TAGTAATTCA TTTAACTCCT CTTACATGAG TAGCGACAAT GAGTCAGATA TCGAAGATGA	60
	AGACTTAAAG TTAGAGCTGC GACGACTACG AGATAAACAT CTCAAAGAGA TTCAGGACCT	120
15	GCAGAGTCGC CAGAAGCATG AAATTGAATC TTTGTATAACC AAACTGGCA AGGTGCCCCC	180
	TGCTGTTATT ATTCCCCCAG CTGCTCCCT TTCAGGGAGA AGACGACGAC CCACTAAAAG	240
20	CAAAGGCAGC AAATCTAGTC GAAGCAGTT TC TTGGGAAT AAAAGCCCCC AGCTTTCAAGG	300
	TAACCTGTCT GGTCAGAGTG CAGCTTCAGT CTTGCACCCC CAGCAGACCC TCCACCCCTCC	360
	TGGCAACATC CCAGAGTCGG GGCAGAATCA GCTGTTACAG CCCCTTAAGC CATCTCCCTC	420
25	CAGTGACAAAC CTCTATTCAAG CCTTCACCAAG TGATGGTGCC ATTTCAAGTAC CAAGCCTTTC	480
	TGCTCCAGGT CAAGGAACCA GCAGCACAAA CACTGTTGGG GCAACAGTGA ACAGCCAAGC	540
	CGCCCAAGCT CAGCCTCTG CCATGACGTC CAGCAGGAAG CCCACATTCA CAGATGACTT	600
30	GCACAAGTTG GTAGACAATT GGGCCCGAGA TGCCATGAAT CTCTCAGGCA GGAGAGGAAG	660
	CAAAGGGCAC ATGAATTATG AGGGCCCTGG AATGGCAAGG AAGTTCTCTG CACCTGGCA	720
35	ACTGTGCATC TCCATGACCT CGAACCTGGG TGGCTCTGCC CCCATCTCTG CAGCATCAGC	780
	TACCTCTCTA GGTCACCTCA CCAAGTCTAT GTGCCCTCCA CAGCAGTATG GCTTTCCAGC	840
	TACCCCATTT GGCGCTCAAT GGAGTGGGAC GGGTGGCCA GCACCACAGC CACTTGGCCA	900
40	GTTCCAACCT GTGGGAACTG CCTCCTTGCA GAATTTCAAC ATCAGCAATT TGCAGAAATC	960
	CATCAGCAAC CCCCCAGGCT CCAACCTGCG GACCACTTAG ACCTAGAGAC ATTAACTGAA	1020
45	TAGATCTGGG GGCAGGAGAT GGAATGCTGA GGGGGTGGGT GGGGGTGGGA AGTAGCCTAT	1080
	ATACTAACTA CTAGTGTGC ATTTAACTGG TTATTTCTTG CCAGAGGGGA ATGTTTTAA	1140
	TACTGCATTG AGCCCTCAGA ATGGAGAGTC TCCCCCGCTC CAGTTATTGG AATGGGAGAG	1200
50	GAAGGAAAGA ACAGCTTTTT TGTCAGGGGG CAGCTTCAGA CCATGCTTTC CTGTTTATCT	1260
	ATACTCAGTA ATGAGGATGA CGCCTAGGAA AGTCTTGTC ATAAGGAAGC TGGAGAACTC	1320
55	AATGTAAAAT CAAACCCATC TGTAATTTCG AGTGGGTGGA GCTCTTGCTT TTGGTACATG	1380
	CCCTGAATCC CTCACCTCCCT CAAGAACCG AACCAACAGGA CAAAAACAC CTACTGGCT	1440
	CTCTCCTTACCC CTGCCCCCTT CCCTTTTTT TACCCCTCTC TTTTTTATTT TTTCTTGCT	1500
60		

CTTTAGAAC CAGTGAAAAA TACCAGGGTA CTGGGGTGC A	1560
ATTAGTGCTT TAAGCAAAAG ATATTAGCAG CTTTGACTGC AGCATTAGCA ATTAGGAAA	1620
5 AAAAAAANWA AAAACTCGAG GGGGGGCCG GTTACCCAAT TCGCCCT	1667

10 (2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1408 base pairs
 - (B) TYPE: nucleic acid
 - 15 (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:	
20 ATTACACACC TGAGCACTGT GCCTGGCAAG ACCTGTCTTA ATAGATTAGA GAACCACTGA	60
TAGATGGTCA GCTTTCTGTA GCAGTGAGAA CCCTACATTT CAAATGIGGA TAGCACCTT	120
25 GCGGGGAAAC ATCACTTGGC ACATCTGCAT TCTTTTTGCA CACAGGGTCT CACTCTGTTG	180
CCCAGGCTAG AGTGCATGGC ACGATCTTAG CTCACTGCAA CCTCCACCTC CCAAGTTCAA	240
GOGATTCTTC TGCCCTCAGCC TCCTGAGCAG CTGGGATCAC AGACATGCGC TACCATGCC	300
30 AGCTAATTTT TTGTATTTTT TGTKKGTTTG TTTTTGTTK TAAGTAGAGA CGGGCTTTCA	360
CCACGGTTGGS CAGGCAGGTC TCGAACTCCT GAMTCAGGT GATCCACCCA CATCTGOGTT	420
CCAATATCTT TCTAACATA ATGATAGCCG TAATTAATAT TTTCCAGTAC ATTTTTATGC	480
35 CTTTACACAC GAGAGTGGTA GACAGACACA AACCCAGATC TGCTGACTC CAAAGCCCGT	540
TTGTCATCAT TCCTTTACG GTATCCTATA GTGGTATCCT TTACAGAAAG ACAGCTTTA	600
40 CCCAACAAAG ACTTAACTTC CCAGGATGCC AGAAGGACAA AGCCGGATG CTTTTAAGRA	660
GRAAGTTATC AAGAMCTTAT TTTATAAATG AGATTAGATA GGGAAAGGCA ATTTATCTTT	720
ATTAAAAACT GAAAAGGCCA GCATAGGGAA GGAGGTCTT CGGTGGTCTT TTTCAGGGAA	780
45 ATACCTCAGT TGCTTTATT AGAAACAGAT AGTACCTAAG GTTTTGAGGT AGGWACAGCT	840
TAAGGCATGC TAATGKTCAT GGGCCTTCC ATAGTCATTT TKGTATTTTG GTTWACATTT	900
50 GAGCAATAGG CAGCCCTTCA CTGCTGCTGG AYTCACTCCT GCCAYTATTA CAGGTGACAG	960
AGGAGACAGG AGGTATGCT TTTCTATTT TAWACATGCT TTATATTAA CACAAGCTCT	1020
TGGGTATCTT AGATAAACAG AAGTTGCCTA GCACCTCTT TAGTCATG AACCCCTTAA	1080
55 CATTAAAGCA AAATAATAAA CAGTCCTTGT AGGTTCCCTA ACAATGAAAC GTGTTGAGT	1140
GGCAGCAGCG GAATCCATGC YTCTTCTCCT GGAGTGTGCA AKAGTCCGTG GTCTGAGTA	1200
60 TCTCACACAG ATGTGGCATT TTATGTGIGA TGCTCTAATT AAGGCCATG GTACAGAAC	1260

AGATTCAGAC GTCCTCTCAG AATAATGCA TTCTTTGCA AAGGTGAATA TTTTCTCTT
1320
AAAAAAATATG TATTAAGGTTG TAGTTTCATT TATTAGTCCT GCTAAAAAAA AAAA
5 AAAA
ACTTNGAGGG GGGGNCGGT ACCCAATT
1380
1408

10

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2031 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

20 AGGATATGCA TGATTCTAA CCAGGCTATA TGTTAAAAAA AAATTGGAAA ATGCAATACA 60
TTTTTAETA TACAAACTAC AGAATGAGTA TGCAAGTTTT ATTATATCAAATGTAATGGA
25 TTTTTAAAGG CTGAGAATT TCCCTTATAC CTACCTTTTC AGTTATTTTA ATTATACCAA 180
ATTATCAACT AGAATAGCTT CATCCATATG AAATATAAAA TGAAGAGACA CCTAGGCTCT
30 ATCAGGCTTA GGATTCTTG AACTTATTTTC CACTTTAATT TCTCAGTGGAGTTAAGAGG 300
GGTGAGAATA CAAAGAGGG GAAAGACTGA CAACTAACAA AACCAGCACC ACATCGCTAG
35 GTGGTGCTTA CTATTACCT TCTCAGGATT TTCCTCAGAT TGAAAAGCTT ATGAGGATT 420
CTTGGGACTC TTATATACCT GCCTGTTAGT ACAGAGCTTT CCTGATGATA TTTACTCTTG
40 ACCACATGTG GTTGTAAAC CTTAACTTTC TTTCTCCAGG AGGGTGGTGA TAGAACAGA 540
GGTAGTATT TATGAATGTA TGTCTCGTG AAATGTGAG GGTGGGGAGA AAGACTTTA
45 ACGGAGGAGA GCCATCTATT TTGTCCTAA AGCCACCTCT CAGCAGAACATC GTCATGTTT 660
TCTGATGCAC CGCTCTCTT CATGCCAAG ATGACTTGCG AGGCAATCTC AGGAGCTGTG
50 GACTTAACCR TTGCAARGCA CACTGTCTTT CTCAGCGTC TCTGCAAGTC AGTAGGTGTT 780
AGTATGGTTG CAAAGTCAC TGTCTCAGCA AAGTTGAACG GGGCTACCTC TCTACAGCTG
55 TTTCTCTAGA GGGAAAATC TTGAGACCAAG ATGGTGGAGC TCTGGAGTCAGA GAGGAAATGG 840
GTTGCTTCAG CACAAAGCTG CTGCTTTAC TTCAGCCACT TCTGACATT TTACATACCG
60 AGCCTGAGAT TGTGTGATTA TCTCAAATCA AATCACTTTG ATGGAGATAA ATAATCAAAA 960
CTGTTTATA GTCATTGATT TGGTGAGAAC AGTAATGGAA AATGGTGTG AAGGACTTCT
65 CATTGTTGGA GCTTCCCTTC CAGAGTCCTG GCTGATGGT GTTCGCTGTT CATCTGAGCC
70 CCCAAAGCA TTATTACTGA TACTTGCACA CAGTCAAAAG CGCAGACTGG ATGGATGGTC 1020
75 1080
80 1140
85 1200

	TTTTATAAGG CATTAAAGGG TACACTACTG TGTTTCACTG ACCATACATT TTTCTTAGCC	1260
	CCTCAAGTAA TATAGCACAG AGTTATGAAT GACAATTCCC CTAACCATTG CTCTTCATAT	1320
5	CTGCCTCTTC CCCTTACCAT CGTAATTCTC CAAACTGGTC ATAAAGGCAC TCTGTGAAGA	1380
	TATTGGGGAC TGACATCTTA AGCTCTCACC TGGCTGCAGT AGGAAAGGCC AACTGACGA	1440
	CAAAAAAAA ATTCTTTATA AAGATGATAT GGTAACATGT ATCTTTGCCG TGGGTCTGGG	1500
10	TGGGTCAGT CAGTCTCAGA TTACAAGCA TTAGGAGCC TAGGAAAAG CTGCTAGTAT	1560
	TCTTTAAAAA GTTACATTAA TGACTTGCAA TGATAGAAAA CTCCCTCCAA TAAATGGCA	1620
15	TTTTATAATA TTATGTGTGT ACTTCACAGT GTAAAAATA CCCTCATACG TTATTGCATT	1680
	TGATCTTCAC AGAAAGTGCA TTAAACCAG TACTCTGGGT GCAATAATA ATATGTAGAA	1740
	ATTTAAGTCC TCCAATTCCA GCATATCCAG TGAGTTTGA CAGTGTGTTT ATGTGGAATG	1800
20	TTTAAGGATA TACAATTGTA CTTTATATAA ATTGGTTCTT GTTCTTCITA AATGTGACAT	1860
	GAAATAATTG TGCTGCTACA TTATACTGGA AATTAACAGG GGAAAGGGAG AGAGCTCTG	1920
25	GCTCCCTTGA GGTTCTGCTA GTGGTGTAG GAGTGGTAC AACTGAGCTT TTAGTAACCA	1980
	TTTAACCGTA TGTAAACTTG GTTCTAATT AAAAATTTTCC A	2031

30

(2) INFORMATION FOR SEQ ID NO: 33:

	(i) SEQUENCE CHARACTERISTICS:
35	(A) LENGTH: 971 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:
	CGCGTCGGAA CTCGGCCCGG GGACATCCAC GGGGGCGGAG TGACACGGGG GAGGGAGAGC 60
	AGTGTCTGC TGGAGCCGAT GCAAAACC ATGCATTCT TATTCAAGATT CATTGTTTC 120
45	TTTATCTGT GGGGCTTTT TACTGCTCAG AGACAAAAGA AAGAGGAGAG CACCGAAGAA 180
	GTGAAAATAG AAGTTTGCA TCGTCCAGAA AACTGCTCTA AGACAGCAA GAAGGGAGAC 240
50	CTACTAAATG CCCATTATGA CGGCTACCTG GCTAAAGACG GCTCGAAATT CTACTGCAGC 300
	CGGACACAAA ATGAAGGCCA CCCCCAATGG TTGTTCTTG GTGGTGGCA AGTCATAAAA 360
	GGCCTAGACA TTGCTATGAC AGATATGTGC CCTGGAGAAA AGCGAAAAGT AGTTATAACC 420
55	CCTTCATTG CATACTGGAAA CGAAGGCTAT GCAGAAGGCC AGATTCCACC GGATGCTACA 480
	TTGATTTTG AGATTGAACG TTATGCTGTG ACCAAAGGAC CACGGAGCAT TGAGACATT 540
60	AAACAAATAG ACATGGACAA TGACAGGCAG CTCTCTAAAG CCGAGATAAA CCTCTACTTG 600

CAAAGGGAAT TTGAAAAGA TGAGAAGCCA CGTGACAAGT CATATCAGGA TGCAGTTTA	660
GAAGATATTT TTAAGAAGAA TGACCATGAT GGTGATGGCT TCATTTCTCC CAAGGAATAC	720
5 AATGTATAACC AACACGATGA ACTATAGCAT ATTTGTATTT CTACTTTTTT TTTTAGCTA	780
TTTACTGTAC TTATGTATA AAACAAAGTC ACTTTCTCC AAGTGTATT TGCTATTTT	840
10 CCCCTATGAG AAGATATTT GATCTCCCCA ATACATTGAT TTTGGTATAA TAAATGTGAG	900
GCTGTTTGC AAACTTAAA AAAAIIWAAA AAAACTSGAG GGGGCCGT ACCCAANTCG	960
CCGNATATGA T	971
15	

20 (2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1792 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAACCCCTT TCTCTGGTA AAGGGTAAGG GGGGGATAA TGTTCACCAC AGGTACGAAA	60
30 TAGTCACCTT AACATTGAGA CCTCTGCCTC ATTGAATTCA GGTTTTTAA GTACTTGAAA	120
CTCTTCAGAT TCTCCTTATT TTAGTTCTT TTTACATTTA TGAAGTAGAA ACCATTGTTT	180
35 TGTAAACTGT TTTGAAAATA AATAGCCTAG TCTCTTATCC TCTTTAGCGT GGATTAAAGG	240
TGAAGTTCTG CAAATGGGAG AGTGTTCACA GTAGATAGCT CAGATTGATT GAACACATTT	300
40 GAGGAAGAGA CTCTGCATG AGATACCAGC ATTTTTACAA ATACTTTTTA TGTACATTCT	360
TTATTTTGTC ATTTTGTCAA CCCTCTCCCC AAGCACATCT TCTTTCTTT TACTATGCT	420
ATGTAGGGAA AAACAAAACA AAAAATTGCA CTTACGGTAC ACTCCAAAA TGTGGGTAAAT	480
45 CCGTGTCTTT CAAAAAACAT TTCTGTTTT TGTGGTGT TGGTCAGTCC ATTGCATAAG	540
TGACAAGTTT GGGTCTTGT GGCACGTATG TATGAACCGG GAGGGGGATG ASAATTGCCT	600
50 GTCCTTCAGT ARGCTGTAAA AGTAATTAC ATGTAAGTAA AAAGGGAAAA TAGAATAGAT	660
GCCAAAGTCA TTTATTCACT CCTTAGTTTT CTTATGTGGC ATTACTGCAT CTGCTAGTTA	720
GTGAGAAAGC ACCCTCAGCT TTTACTGCTC CCCTCCCTGC CTGCCAACAC ACTTGATGTG	780
55 TGCAAACAGC CCTCAAGTAT CTGTCAGATG ACCTATATAA GGTATTGAAT AAGGTATTCT	840
TGTCAGTTA GAAATGGACT GGATAAAACT TACTMGGTIG TCATTATTTT ATCTCATTIG	900
60 TCCTGTTACA TGCCCTATGT TAAGATAATT ATATTGCCAC TAATAATCAA GATGCTAAAT	960

	GAGTATTACA ACTGGCTAAT ATCATTTTT ATATAACAAGG GTATGTGTAT ATTTGGAATT	1020
	GRTATGAGAA ACTCATTGT ACCCATTTGA GTGATATTGC ACAACAAACA CAGATAYCTA	1080
5	CAGACTCCGT TTTCATTTTC TCGTGTCTT TATGATAATG ATCTTGTAG ATTGGTTATT	1140
	TCTGTACTTT ATCTGTAATA AACCTGTAG ATCCTGTGAA CCATTACTTT GCCTAAATCA	1200
10	CTTGAGACTT GAGTCTTTAA TAACAAAGCA TCAATATTCA CTAAAGCAA TCTCTTTGA	1260
	GTTTCTGTGA CTTGGCTAGA AGCTCTTGAC ACTAAGGGAT TAGTGTAAAT TTTCCCTGGG	1320
	GGTGTCCAC TAGGGCATT A CTGTATAATG ACTTGATGTT GCCACATAGA CTTCAAGATA	1380
15	TATAATATT TGAGGATTTT GTGATGGC CTATGTTTA TTGCATAGTG TGAAACGTGT	1440
	AAAGCTTGGT TAACCTGTAT ATAGATAGCT TATTGTGAC TAGTTATAGT GTATTTAGGG	1500
20	TTGCCIGTAA TATTTAAGCT TCTTACTGTA TGTGTGCT GGTAGGAACA TATAATT	1560
	GTACATTATA TTTACTGAGA TGTTGCCCTT TTTATTTAC AAATACTTTG GAATTCCAAT	1620
	GTGTTTTTG CTTCCGTGAG GATTAATTG GAAAGGTTT TAATGACATT CCACTGATT	1680
25	CAGATTTGC TTGAGATTGA CTTCAATAAA TTGTCCGTAA TGTTCCAAAA AAAAATTAAA	1740
	AAACTCGAGG GGGGCCGGT ACCCAANNCG CGGGATATGA TCGTAAACAA TC	1792

30

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 896 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

	AGTTGNANAC AACAGGACCT GAGTCCTTGG GCAGCACCAAG TAGGTGCCCC CYTGCYTCYT	60
	GCCAGCYTCA CYTGCCACYT TYTGCCCTY TCGGGATGCC TTGCGCAGACA GAGYTYTTCG	120
45	CTGCCTGTGG TGGCCAYCT TTGCTTTGG TTTCCTTGC CTTGGCCTC CCTTTTGTIC	180
	CCCGGGCAGC CTTGTGTGAC CTGCCCTTT CCCTCCCTTC CTTTCCAGGA CAAGCACGCC	240
50	GAGGAGGTGC GGAAAAACAA GGAGCTGAAG GAAGAGGCCT CCAGGTAAAG CCTAGAGGCC	300
	AAAGAACTTT CCAGGTCAAGC CGGACAGCTC CAGCAGCTCC ACGTTCCAGG CAGCCTCGMC	360
	CGCCGGCTGC GCTCCCAGCA CTGGGGTTTG GGGGGAGGGG GGTGCCAAG GGGCGTTTCC	420
55	TCTGCTTTTG GTGTTGTAC ATGTTAAGAA TTGACCAGTG AAGCCATCCT ATTTGTTCC	480
	GGGAAACAAT GACGGGGTGG GARAGGGAG AGGAGAGAGT TTGGAAAGG GAGATGGAGA	540
60	AGAACTCAAG GACATTGCAA CCTGGCCGG CGCAGATCTG ATTTTCACAT CTCTACCTGG	600

ACATTGAGCC TCCCAGGCAC CATGTGAGG AGAGATGAAA ACCAGGGCGG TAGAACTTCA	660
GGGTGAAGGA CAGAGTCCTG GGTGGGGCAG CGGCTGCAGG GCGCACCAAGA GAACCCAGCC	720
5 AGAGGGGGTG TGAGTACCAAG TGGTGTGCT TCCACCCCTGC AGCAGGTGGG ATGAGGTCTG	780
TGTGTGTG TGAAACCATCA TTTTTGATC ATCATGACCA ATGAAACATT GAAAAAAA	840
10 AAAAAAACTG GAGGGGGGCC CGTACCCAAN TCGCCGNATA GTGATCGTAA ACAATC	896

15 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 912 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

25 TCGACCCACG CGTCCGGTCA GCCAGTCGCA TCCAGCCATG ACAGCCTTCT GCTCCCTGCT	60
CCTGCAAGCG CAGAGCCTCC TACCCAGGAC CATGGCAGCC CCCCAGGACA GCCTCAGACC	120
30 AGGGGAGGAA GACGAAGGGA TGCAGCTGCT ACAGACAAAG GACTCCATOG CCAAGGGAGC	180
TAGGCCCCGGG GCCAKCCGCG GCAGGGCTCG CTGGGTCTG GCCTACACGC TGCTGCACAA	240
35 CCCAACCTTG CAGGTCTTCC GCAAGACGGC CCTGTTGGGT GCCAATGGTG CCCAGCOCTG	300
ARGGCAGGGA AKGTCAACCC ACCTGCCCCAT CTGTGCTGAG GCATGTTCTT GCCTACCATC	360
CTCCTCCCTC CCGGGCTCTC CTCCCCAGCAT CACACCAGCC ATGCAGCCAG CAGGTCTCC	420
40 GGATCACYGT GGTTKGGTGG AGGTCTGCT GCACTGGGAG CCTCARGARG GCTCTGCTCC	480
ACCCACTTGG CTATGGGAGA GCCAGCAGGG GTTCTGGAGA AAAAAACTGG TGGGTTAGGG	540
CCTTGGTCCA GGAGCCAGTT GACCCAGGGC AGCCACATCC AGGGTCTCC CTACCCCTGGC	600
45 TCTGCCATCA GCCTTGAAGG GCCTGATGA AGCCTTCTCT GGAACCACTC CAGCCCAGCT	660
CCACCTCAGC CTTGGCCCTTC ACGCTGTGGA AGCAGCCAAG GCACCTCCTC ACCCCYTCAG	720
50 CCCCACGGAC CTYTGTGGGG AGTGGCCGGGA AAGCTCCSG GCCTYTGGCC TGCAGGGCAG	780
CCCAAGTCAT GACTCAGACC AGGTCCCACA CTGACCTGCC CACACTCGAG AGCCAGATAT	840
TTTGTAGTT TTTATKCCCT TGGCTATTAT GAAAGAGGTT AGTGTGTTCC CTGCAATAAA	900
55 CTTGTTCTG AG	912

60 (2) INFORMATION FOR SEQ ID NO: 37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1382 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

10	AATTCGGCAC GACCGGAGGC GAGGGAAACT RAGGGGAAA GTTGTGTGTC GTGTTGGCAG GAGGGCCTAG AAGGGAAAGA CTGTCTAGTG GGACAATGTC ATATTATAAA TTTGGAATGC	60 120
15	TGAATAGAAA ATTATAGATT TTGATATTGA AGGAAATGAA GCGAAGCYTA AATGAAAAATT CAGCTCGAAG TACACGAGGC TGTTTGCCTG TTCCGTGTT CAATCAGAAA AAGAGGAACA	180 240
20	GACAGCCATT AACTCTAAT CCACTTAAAG ATGATTCAAG TATCAGTACC CCTTCTGACA ATTATGATT TCCTCCTCTA CCTACAGATT GGGCCTGGGA AGCTGTGAAT CCAGAGTTKG	300 360
25	CTCCGTGTAAT GAAAACAGTG GACACCGGGC AAATACCAACA TTCAGTTCT CGTCCTCTGA GAAGTCAAGA TTCTGTCTTT AACTCTATTTC AATCAAATAC TGGAAAGAAC CAGGGTGGTT	420 480
30	GGAGCTACAG AGATGGTAAC AAAAATACCA GCTTGAAAAC TTGGRATAAA AATGATTTTA AGCCTCAATG TAAACGAACA AACTTAGTGG CAAATGATGG AAAAAAATTCT TGTCCAATGA	540 600
35	GTTCGGGAGC TCAACAAACAA AAACAATTAA GAACACCTGA ACCTCCTAAC TTATCTCGCA ACAAAGAAAC CGAGCTACTC AGACAAACAC ATTCACTAAA AATATCTGGC TGCACAATGA	660 720
40	GAGGGCTAGA CAAAAACAGT GCACTACAGA CACTTAAGCC CAATTTCAGA CAAAATCAAT ATAAGANACA AATGTTGGAT GATATTCCAG AAGACAACAC CCTGAAGGAA ACCTCATTTG	780 840
45	ATCAGTTACA GTTTAAGGAA AAAGCTAGTT CTTTAAGAAT TATTTCTGCA GTTATTGAAA GCATGAAGTA TTGGCGTGAA CATGCACAGA AAACTGTACT TCTTTTGAA GTATTAGCTG	900 960
50	TTCTTGATTC AGCTGTTACA CCTGGCCCAT ATTATCGAA GACTTTCTT ATGAGGGATG GGAAAAATAC TCTGCCCTGT GTCTTTATG AAATCGATCG TGAACCTCCG AGACTGATTA	1020 1080
55	GAGGCCGAGT TCATAGATGT GTTGGCAACT ATGACCAGAA AAAGAACATT TTCCAATGTG TTTCTGTCAG ACCGGCGTCT GTTTCIGAGC AAAAAACTTT CCAGGCATTG GTCAAATTC CAGATGTTGA GATCCAGTAT TATATTAATG TGATGAATGA AACTTAAGTA GTGATAAAAG GAAGTTTAGC ATAAATTATA GCAGTTTCT GTTATTGCTT AATTTACCAT CTCCATAGTT TTATAGCTAC TATTGTATT CACTTGTGA ATTAAAGTAT TTGAATTCTT TTAAAAAAA AA	1140 1200 1260 1320 1380 1382

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 872 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

10	GGGCTACTTC AAAGCCCTGG GCCTTATTTC TTCAGGTAAA AAAATATAAA GTCAGATCTC	60
	ATCCCGGCTG GCCATGCTGT TAGACCCCTTT CATCCTCTC TTCTGCCTCT TCTCAACAGC	120
15	TGCCCGAGTCC TGTTTGAAT TCATATACAT ACAGTTCTAA TACTGATGTA TTTACCCCTCA	180
	TAAGCCACTC AACCCAGAAT CTTATTTGAA TTATAATCCA GAAACATCAG GTGACGTGIG	240
20	AGACTACTGT ATGAGAAAGA GACAGTTAA GGGTCAGTCC AATGGAAAAA AGAGTTCTCA	300
	GAGCTTTCTT TAGCTTATTTC TCATCAAAGA GCTTTCTCTG CAGAAGGAAC CTACTGGTTC	360
	CTCCCTTCCA GTCCTAGAAA TCCTGACCTA GAGTGGCTTA ATCCTGCTAG CACCTCTCTC	420
25	TOCCACTCTG GTGCCAAATG ACTCCAGGAA CTGGGCCATG ATGTTGGGGG AATGACCTTA	480
	CCCTGAGCAT GTCACTCATG CAITGAACAA CAGCTAAGAG CAGAGCTTAG AGCTTAGAGC	540
	TGGGCCCTGT AAGGTGAGAG GAATCACATC CTGCAGAAGT CTGTCCTGAG AAGCAGGTAC	600
30	TCCTGTCACA GCAGAGACAC AGTGGATACC TGAGTAACAA TAATACAAGA CAGGACGTGG	660
	GMACAGCAAA AGATTTGGGT GTCAGAAGAR CCCGAGAACCA CTTYCAGGCA GGAACATTCA	720
35	RARTTGTCT TGGAGGAART AGGCMCSAAG GCTGGCAGG ATTCMCGGG GCAGAGATGG	780
	AGCAAGCAAT TGAAATGAAA GOCATGGCAT GGGAAAAGGA GCACIIGGCCA CAGGGAGTGC	840
40	AACGTTGTGA TCGAAGGCCA CTGTGGAGCC AT	872

(2) INFORMATION FOR SEQ ID NO: 39:

45	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 812 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
50	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

55	GGCAGAGGCT CACCCCAGCA GAGATTGAGG GGGAACCGTG ATGAAATTTC TAAGTATTCT	60
	GCTTGATGAT AATAATTTCCTCTTATGTT AATGTTGGCT CCGTTGGGT GTTTAGCTTT	120
	TGAAAGGAGT ATGAAAATGC GGAATGGGGC TTTGGGCCTT GAGGAGGTGT GATCTCTAGT	180
60	GTTTAAAAAA TTAAATTGCA CAAATAGAAA TAATTCAACCC ACATTATTGA ACCCCACTAA	240

AGCATATCCT TTTTGTCCAT ATTCCCTTCG TGCTGCCCTC GTGTGTACCA TTATTACTCA	300
5 GTTGTGATTG GAGCTCGTTC CACTAAAGT CATTAGA TACTTTGGG TCGTGTGKGA	360
ATATTATTC AATTCTATT CTGTGTTTA CTTAATTACT TTATTATGGA ACCTTTACAC	420
10 AGGTCTGGTG TACTTGTCT TTGAAAAGTC TTATGTTGAC CACCATCACT GAGCATATAG	480
CTTTTCCTT ATTCCTTGG GATAATTACC CGAAGTGGAA ATACCGAAC AAACCTCTGT	540
TTTCTTCTT TGGCACTATT ATATAAATTG TTTCCAAAC AAGGCATGTT TACAATAGAC	600
15 ATTTTCAAA ATCTGGGTAT TTGCTCTATT TTGCTCTCTG TATGCAGAAT TCAGCGGGGT	660
GCCAAGTCGT TTTCTGTGTG GGTTGAGAGA CAGGCTGTGC AGCCCACITGT TGCATAGGAC	720
TAACTACTAC AAATCATGCT GAGACCGAGC TATTTTGCT GCTTAGARGC TTTGCAGCCT	780
20 TGAGTAAGTT TCGNCATCTG GAAACNTTGN AA	812

25 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1515 base pairs
- (B) TYPE: nucleic acid
- 30 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

35 AATTGGCAC GAGGAAATT CAAGCACTTT TCCTAAAAGA AGGGGGAATG GATGCTGAAA	60
CAACACGTNT CCCACAAAGG GAGCAGACAC TGGGCTTG TG AAGCTGCCCT ATACCTTC	120
40 CACAGAACTG GGGTCCCCC TCCCTGACAT GCAGATTTC ACCCAGAAGA CAGAGAAGGA	180
GCCAGTGGTC ATGGAATGGG CTGGGTCAA AGACTGGGTG CCTGGGAGCT GAGGCAGCCA	240
CCGTTTCAGC CTGGCCAGCC CTCTGGACCC CGAGGTTGGA CCCTACTGTG ACACACCTAC	300
45 CATGCGGACA CTCTTCAACC TCCCTGGCT TGCCCTGGCC TGCAGCCCTG TTCACACTAC	360
CCTGTCAAAG TCAGATGCCA AAAAAGCCGC CTCAAAGACG CTGCTGGAGA AGAGTCAGTT	420
TTCAGATAAG COGGTGCAAG ACCGGGGTTT GGTGGTGACG GACCTCAAAG CTGAGAGTGT	480
50 CGTTCTTGAG CATGCGAGCT ACTGCTCGGC AAAGGCCCG GACAGACACT TTGCTGGGA	540
TGTACTGGGC TATGTCACTC CATGGAACAG CCATGGCTAC GATGTCACCA AGGTCTTGG	600
55 GAGCAAGTTC ACACAGATCT CACCCGTCTG GCTGCAGCTG AAGAGACGTG GCCGTGAGAT	660
GTTTGAGGTC ACGGGCTCC ACCGACGTGGA CCAAGGGTGG ATGCGAGCTG TCAGGAAGCA	720
60 TGCCAAGGGC CTGCACATAG TGCCCTGGCT CCTGTTGAG GACTGGACTT ACGATGATT	780

300

	CCGGAACGTC TTAGACAGTG AGGATGAGAT AGAGGAGCTG AGCAAGACCG TGGTCCAGGT	840
	GCCAAAGAAC CAGCATTTCG ATGGCTTCGT GGTGGAGGTC TGGAACCAGC TGCTAAGCCA	900
5	GAAGCGCGTG ACCGACCAGC TGGGCATGTT CACGCACAAG GAGTTGAGC AGCTGGCCCC	960
	CGTGCTGGAT GGTTTCAGCC TCATGACCTA CGACTACTCT ACAGCGCATC AGCCTGGCCC	1020
10	TAATGCACCC CTGTCCCTGGG TTCGAGCCTG CGTCCAGGTC CTGGACCCGA AGTCCAAGTG	1080
	GCGAAGCAAA ATCCTCCTGG GGCTCAACTT CTATGGTATG GACTACGCGA CCTCCAAGGA	1140
	TGCCCGTGAG CCTGTTGTCG GGGCCAGGTA CATCCAGACA CTGAAGGACC ACAGGGCCCCG	1200
15	GATGGTGTGG GACAGCCAGG YCTCAGAGCA CTTCTTCGAG TACAAGAAGA CCCGCAGTGG	1260
	GAGGCACGTC GTCTTCTACC CAACCCCTGAA GTCCCTGCAG GTGCGGCTGG AGCTGGCCCC	1320
20	GGAGCTGGGC GTTGGGGTCT CTATCTGGGA GCTGGGCCAG GGCTGGACT ACTTCTACGA	1380
	CCTGCTCTAG GTGGGCCATTG CGGCCTCCGC GGTGGACGTG TTCTTTCTA AGCCATGGAG	1440
	TGAGTGAGCA GGTGTGAAAT ACAGGCCTTC ACTCCGTTAA AAAAAAAAAA AAAAAAAAAA	1500
25	AAAAAAAAAA AAAAA	1515

30 (2) INFORMATION FOR SEQ ID NO: 41:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 704 base pairs	
	(B) TYPE: nucleic acid	
35	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:	
40	AAGATGGTGG CGCCCAGAGC TTCCCTCTAT GCTGCTCCCC TGAGAGAGGC GTTCCATCA	60
	ACCAAGTTTG CAAGGAGTTC AATGAGAGGA CAAAGGACAT CAAGGAAGGC ATTCCCTCTGC	120
45	CTACCAAGAT TTTAGTGAAG CCTGACAGGA CATTGAAAT TAAGATTGGA CAGCCCACIG	180
	TTCCCTACTT CCTGAAGGCA GCACCTGGGA TTGAAAAAGGG GGCCCGGCAA ACAGGGAAAG	240
	AGGTGGCAGG CCTGGTGACC TTGAAGCATG TGTATGAGAT TGCCCGCATC AAAGCTCAGG	300
50	ATGAGGCATT TGCCCTGCAG GATGTACCCC TGTCGTCGTG TGTCCGCTCC ATCATCGGT	360
	CTGCCCGTTC TCTGGGCATT CGCGTGGTGA AGGACCTCAG TTCAGAAGAG CTTGCAGCTT	420
55	TCCAGAAGGA ACGAGCCATC TTCCCTGGCTG CTCAGAAGGA GGCAGATTIG GCTGCCCAAG	480
	AAGAAGCTGC CAAGAAGTGA CCCTTGCCCC ACCAACTCCC AGATTCAAA GGAGGTAGTT	540
	GCAAAAGCTG TGCCCAAGGG GAGGAAGGAG GTCACACCAA TATGATGATG GTTTTCATGA	600
60	CTTTGAATGA TATATTTTG TACATCTAGC TGTATGGAGG CATCAGGCCT GAATAAACAT	660

CCTTCTTAA AAAAAAAA AAAAAAAA AAAAAAAA AAAA 704

5

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1094 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGCAGCTTTC TTACAAACCC ATCCTCTGA AATGTTGCTT CAAATTCA	60
20 CAGCCCCACT ATTCCACACA TACTGTTACT GTTTCTTAT CCTACTTTCT CAATTTGGA	120
ACATAGTTGC AGTTACTGCA TTGAATACCT GTGGGTTGC CTGTTGTTCT GTCTGTCT	180
25 GTGGTTCTTG TAATANTGGA TCCCAGAGAT AAAATGGACA GTTGTNATGC ACAGTTAATT	240
CAGAAACTAG ACCTTACTTG CTGTGTGAAA TACCAACTAA ATTCTCAGTG AACTCAGCTG	300
30 ANCTTTATCT CCTTTTGTIT CCCCAATTAA TAATTTCACT TCAGGCCAG AAAGATGGAA	360
TOCCAGCTAA GAAATACAAG TTACACCTG TACTAGCAGC CCATGIGTGC ATGTTCTTAA	420
AGTGCTCTTG CAGCTATGTC ATTATATATTG ATTTCCTGT ATTATTATAA GCAAAGCAA	480
55 TTTGAGGAAA AAAACCCATA ATACCACACC TCATTTTTTT CAAGTAATAG GGTCTAAAGT	540
CTCATYCTYC ATATAATATG TTGAGTATGC AGTATATTAT GTGTTAGGCT CTGGANAGGC	600
AGAGGTTAGA TCATGTWACA GATQATATCK GATTAGGCAG ATAAACAGTA TTTAACCTT	660
60 TTCCTTATTA TATGTAACCTT GCTTCAGGT TTTTAATGT TACTATTATG TCTTTAATAT	720
ATTATCTTTA TTTGTACTTT TGTATACAGA GTGATTTCC TTTTTAAAA AAAATTGTGT	780
CTTTAGGATG GATTCCAAG ATGTGGAATC AGTAGGTTA AGGAATATGG ATATTTGGC	840
75 TGGCAACGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAGGTG GGTGGATCAC	900
CTGAAGTCAG GAGTTCGAGA CCAGCCTGAC CAACATGGCG AAACCCGTGT TNTACTAAAG	960
80 ACACACWWAA AATTRGCCAG TGGTGGTGGC ATGTGCTTGT AGTCCCACTT AGCTACTCGA	1020
GAGGCTGAGG CAGGAGAATC GCTTGAACCC GGGAGGCAGA GGTTGCAGTG AGGCAAGATG	1080
85 GCACCTCTAC ACTC	1094

55

(2) INFORMATION FOR SEQ ID NO: 43:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

TGGCTTGGC	CACGCCCTT	CCCTGGCTG	GAACTACTGG	ACAGACCCTT	TTGAGATGTG	60
10 CCTGTGGTC	TGTGGAGATG	TGTGTAGTGG	TCTTAGCTCT	TTGTTGAGCT	TGTGTGTGTG	120
TTGTGTACTC	TTAGCTGTAT	GCTGAAATTG	GGCGTGTGTT	GGAGGGCTTC	TTAGCTCTTT	180
15 GGTGAGATTG	TATTTCTATG	TGTTTGTATC	ASCTGAATGT	TGCTGGAAAT	AAAACCTTGG	240
TTTGTGAGG	CTCTTTTG	TGGGAACTAA	GTAGGGAAA	AGGTCTTTGA	GGGTTCTTAG	300
30 GCTCCTTCT	ACAAACAGGAA	AATGCCTCAA	AGCCTTGCTT	CCCAGCAACC	TGGGGCTGGT	360
20 TCCCAGTCCC	TGGCCTGCC	CCCTCCTGGT	TCTTATCTCA	AGGCAGAGCT	TCTGAATTTC	420
AGGCCTTCAT	TCCAGAGCCC	TCTTGTGGCC	AGGCCTTCCT	TTGCTGGAGG	AAGGTACACA	480
25 GGGTGAAGCT	GTGCTGTAC	TTGGGGGATC	TCCTTGGCCT	GTCACACCAA	GTGAGAGAAG	540
GTACTTACTC	TTGTAACCTCC	TGTTCAAGCCA	GGTGCATTAA	CAGACCTCCC	TACAGCTGTA	600
GGAACTACTG	TCCCAGAGCT	GGGGCAAGGG	GATTTCTCAG	GTCATTGGA	GAACAAGTGC	660
30 TTTAGTAGTA	GTCTTAAAGTA	GTAACTGCTA	CTGTATTTAG	TGGGGTGGAA	TTCAGAAAGAA	720
ATTTGAAGAC	CGAGTCATGG	GTGGTCTGCA	TGTGAATGAA	CAGGAATGAG	CCGGACAGCC	780
35 TGGCTGTCT	TGCTTTCTTC	GTCCCCATTT	GGACCCCTCT	CTGCCCTTAC	ATTTTTGTTT	840
CTCCATCTAC	CACCATCCAC	CAGTCTATTT	ATTAACCTAG	CAAGAGGACA	AGTAAAGGGC	900
CCTCTGGCT	TGAATTTGCT	TCTTCTTTC	TGTGGAGGAT	ATACTAAGTG	CGACTTTGCC	960
40 CTAACTTAIT	TGGAAATCCC	TAACAGAATT	GAGTTTCTA	TTAAGGATCC	AAAAAGAAAA	1020
ACAAATGCT	ATGAGGCCA	TCAGTCAGG	GTCACATGCC	AATAAACAAAT	AAATTTTCCA	1080
45 GAAGAAATGA	ATOCATCTA	GACAAATAAA	GTAGAGCTTA	TGAAATGGTT	CAGTAAGGAT	1140
GAGTTGTTG	TTCTTGTGTT	TGTTTTGTTA	AAGACGGAGT	CTCGCTCTGT		1200
CACTCAGGCT	GGAGTGGCAGT	GGTATGATCT	TGGCTCACTG	TAACCTCCGC	CTCCGGGTT	1260
50 CAAGCCATTC	TCCTGCTCA	GTCTCCTGAG	TAGCTGGAT	TACAGGTGCG	TGCCACCATG	1320
CCTGGCTATT	TTTGTGTTT	TTAGTAGAGA	CAGGGTTCA	CCATGTTGGT	CGGGCTGGTC	1380
55 TCAAACTCT	GACCTCTTGA	TCCGCCGTGCC	TGGGCTCCC	AAAGTGTGAG	GATTACAGAT	1440
GTGAGGCCAC	CGTGCCTAG	CCAAGGATGA	GATTTTTAAA	GTATGTTCA	GTTCGTGTGTC	1500
ATGGTTGGAA	GACAGAGTAG	GAACGATATG	GAAAAGGTCA	TGGGAAGCA	GAGGTGATTG	1560
60 ATGGCTCTGT	GAATTGAGG	TGAATGGTTC	CTTATTGTCT	AGGCCACTTG	TGAAGAAATAT	1620

GAGTCAGTTA TTGCCAGCCT TGGAAATTAC TCTCTAGCT TACAATGGAC CTTTGAACT	1680
5 GGAAAACACC TTGCTCGAT TCACITTAAC ATGTCAAAAC TAATTTTAT AATAATGTT	1740
TATTTTCACA TTGAAAAAAA AAAAAAATT AAAAACYCGG GGGGGGCCS G:ACCCCAT	1800
10 NGCCCCTAAG GGGGGGGTT T	1821

10

(2) INFORMATION FOR SEQ ID NO: 44:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1024 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGGCACAGT TGAAGAAGCG ACCGAGGGAC TGGGAGTCGT TAGTGAGGAT GACCGGGCAT	60
25 GGCAAGAACT GCACCGCAGG GCCGCTACA CCTACCACGA GAAGAAGAAG GACACACCG	120
CCTCGGGCTA TGGGACCCAG AACATTOGAC TGAGCCGGGA TGCCGTGAAG G:CTTCGACT	180
30 GCTGTGTCT CTCCCTGCAG CCTTGCCACG ATCCCTGTGT CACCCAGAT GGCTACCTGT	240
ATGAGCGTGA GGCCATCCTG GAGTACATTG TGCACCAGAA GAAGGAGATT GCCCGGCAGA	300
TGAAGGCCTA CGAGAAGCAG CGGGGCACCC GGCGCGAGGA GCAGAAGGAG CTTCAAGGGG	360
35 CGGCCTCGCA GGACCATGTG CGGGGCTTCC TGGAGAAGGA GTCGGCTATC GTGAGCGGC	420
CCCTCAACCC TTTCACAGCC AAGGCCCTCT CGGGCACCAG CCCAGATGAT GTCCAACCTG	480
40 GGCCCAGTGT GGGTCTCCA AGTAAGGACA AGGACAAAGT GCTCCCCAGC TTCTGGATCC	540
CGTCGCTGAC GCGCGAAGCC AAGGCCACCA AGCTGGAGAA GCCGTCCCGC ACGGTGACCT	600
GCCCCATGTC AGGGAAGCCC CTGCGCATGT CGGACCTGAC GCGCGTCAC TTACACACCGC	660
45 TAGACAGCTC CGTGGACCGC GTGGGGCTCA TCACCCGCAG CGAGCGCTAC GTGTGTGCCG	720
TGACCCGCGA CAGCCTGAGC AACGCCACCC CCTGCGCTGT GCTGGGCCCG TCTGGGGCTG	780
50 TGGTCACCCCT CGAATGCGTG GAGAAGCTGA TTCGGAAGGA CATGGTGGAC CCTGTGACTG	840
GAGACAAACT CACAGACCGC GACATCATCG TGCTGCAGCG GGGCGGTACC GSTTCGCGGG	900
CTCCGGAGTG AAGCTGCAAG CGGAGAAATC ACGGCCGGTG ATCCAGGCCT GAGTGTGTC	960
55 GGGAGACCAA ATAAACCGGC TTGGGTGCGC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1020
AAAAA	1024

60

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 983 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CGACACGGCT	CGGAGAAGAC	GACAGAAGGG	CCCGACCGCG	AGCCGTCCAG	GTCTCAGTGC	60	
TGTGCC	CCCC	CCAGAGCCTA	GAGGATGTTT	CATGGGATCC	CAGCCACGCC	120	
15	GGCC	CTGGGA	ACAAGCCGGA	GCTGTATGAG	GAAGTGAAGT	TGTACAAGAA	180
	AGGGAGAAGT	ACGACAACAT	GGCAGAGCTG	TTTGCCTGG	TGAACACAAT	GCAAGCCCTG	240
20	GAGAAGGCT	ACATCAAGGA	CTGTGTCTCC	CCCAGCGAGT	ACACTGCAGC	CTGCTCCCGG	300
	CTCC	GGTCC	AATACAAAGC	TGCC	TCAAG	AGG GCTT	360
25	GAC	GAATTCT	GGCGCAAGTT	CCGC	CTGGAC	TGCCC	420
	GGCGG	GCCTTGAC	TGCCC	GTGG	CCATGGAGCG	GATCAAGGAG	480
30	GC	ACCATCAA	GGAC	GACAAG	GGCAACCTCA	ACCGCTGCAT	540
	GT	CTGCTCT	TCATCACGGT	CATGGACAAG	CTGCGCTGG	AGATCCGCGC	600
35	ATCC	AGCCCG	ACCTGCGAGA	GCTGATGGAG	ACCATGCACC	GCATGAGCCA	660
	GACT	TTGAGG	GCCGCCAGAC	GGTCAGCCAG	TGGCTGCAGA	CCCTGAGCGG	720
40	TCAGATGAGC	TGGACGACTC	ACAGGTGCGT	CAGATGCTGT	TCGACCTGGA	GTCAGCCTAC	780
	AACG	CCCTCA	ACCGCTTCCT	GCATGCTGA	CCCCGGGCA	CTAGCCCTTG	840
45	CAGAGTCTGA	GGCGATGGCT	CCTGGTCCCC	TGTCCGCCAC	ACAGGCCGTG	GTCATCCACA	900
	CAACTCACTG	TCTGCAGCTG	CCTGCTGGT	GTCTGCTTT	GGTGTAGAA	CTTTGGGCC	960
	GGG	CCCCCTCC	CCACAATAAA	GATGCTCTCC	GACCTTCAAA	AAAAAAAAAA	983
	KGS	GGCCGGT	CCCC	ANT	CCC	CCC	

50 (2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 2421 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCGGCTGATC	GCTGCCGCTC	CGCCAATACA	ATAGAGCCAK	CCACTACCAG	CAGCCTGGCC	60
------------	------------	------------	------------	------------	------------	----

	CCTTTCTCC TTCTCCAGAG AGACCAATCC AGCCGAACTC GGGGTTGCC TGAGGAGAAG	120
	GAGGAAGTGA CCATGGACAC AAGTGAAAAC AGACCTGAAA ATGATGTCC AGAACCTCCC	180
5	ATGCCTATTG CAGACCAAGT CAGCAATGAT GACCGCCCGG AGGGCAGTGT TGAAGATGAG	240
	GAGAAGAAAG AGAGCTCGCT CCCAAATCA TTCAAGAGGA AGATCTCCGT TGTCTCAGCT	300
10	ACCAAGGGGG TGCCAGCTGG AAACAGTGAC ACAGAGGGGG GCCAGCTGG TCGGAAACGA	360
	CGCTGGGAG CCAGCACAGC CACCACACAG AAGAAACCTT CCATCAGTAT CACCACTGAA	420
	TCACTAAAGA GCCTCATCCC CGACATCAA AAACTGGCGG GGCAGGAGGC TGTGTGGAT	480
15	CTTCATGCTG ATGACTCTCG CATCTCTGAG GATGAGACAG AGCGTAATGG CGATGATGGG	540
	ACCCATGACA AGGGGCTGAA AATATGCCGG ACAGTCACTC AGGTAGTACC TGCAGAGGGC	600
20	CAGGAGAATG GGCAAGAGGGA AGAAGAGGAA GAAGAGAAGG AACCTGAAGC AGAACCTCCT	660
	GTACCTCCCC AGGTGTCAGT AGAGGTGGCC TTGCCCCAC CTGCAGAGCA TGAAGTAAAG	720
	AAAGTGACTT TAGGAGATAAC CTTAACTCGA CGTTCATTA GCCAGCAGAA GTCCGGAGTT	780
25	TCCATTACCA TTGATGACCC AGTCCGAACG GCCCAGGTGC CCTCCCCACC CCGGGGCAAG	840
	ATTAGCAACA TTGTCCATAT CTCCAATTG GTCCGTCCTT TCACTTTAGG CCAGCTAAAG	900
30	GAGTTGGTGG GGCGCACAGG AACCTGGTG GAAGAGGCCT TCTGGATTGA CAAGATCAA	960
	TCTCATGCT TTGTAACGTA CTCAACAGTA GAGGAAGCTG TTGCCACCCG CACAGCTCTG	1020
	CACGGGGTCA AATGGCCCCA GTCCAATCCC AAATTCTTT GTGCTGACTA TGCGAGCAA	1080
35	GATGACCTGG ATTATCACCG AGGCCTCTTG GTGGACCGTC CCTCTGAAAC TAAGACAGAG	1140
	GAGCAGGGAA TACCACGGCC CCTGCACCCC CCACCCCCAC CCCCGGTCCA GCCACACAG	1200
40	CACCCCCGGG CAGAGCAGCG CGAGCAGGAA CGGGCAGTGC GGGAACAGTG GGCAGAACGG	1260
	GAACGGAAA TGGAGCGGGG GGAGCGGACT CGATCAGAGC GTGAATGGGA TCGGGACAAA	1320
	GTTCGAGAAG GGCCCCGTC CCGATCAAGG TCCCGTRACC GCGGGCGAA GGAACGTGCG	1380
45	AAGTCTAAAG AAAAGAAGAG TGAGAAGAAA GAGAAAGCCC AGGAGGAACC ACCTGCCAAG	1440
	CTGCTGGATG ACCTTTTCG AAAGACCAAG GCAGCTCCCT GCATCTATTG GCTCCCACTG	1500
	ACTGACAGCC AGATCGTCA GAAAGAGGCA GAGCGGGCCG AACGGGCCAA GGAGCGGGAG	1560
50	AAGCGGGAA AGGAGCAAGA AGAAGAAGAG CAAAGGAGC GGGAGAAGGA ACCCGAGCGG	1620
	GAACGGAAACC GACAGCTGGA GCGAGAGAAA CGTCGGAGC ACAGTCGGGA GAGGGACAGG	1680
55	GAGAGAGAGA GAGAAAGGGA GCGGGACAGG GGGGACCGAG ATCGGGATAG GGAAAGGGAC	1740
	CGAGAACGAG GCAGGAAAG GGATGCCAGG GACACCAAGC GCCACAGCAG AAGCGGGAGT	1800
60	CGGAGCACAC CTGTGGGGGA CGGGGGTGGG CGCGCTAGC TGGGAAAAACA CTAGAGCTGC	1860

	AGGTACCAGC CACTCGGCC CAGGGGTTA TGGCCACAGA GGGATAGGCA CAGTCTCCAC	1920
	CACCCCTGGAG CCAAGGGTCT TTCACATCAC CTATCCCTAC ATACATACCA AATGGAAAG	1930
5	TGGCCATCCT TTTCCCCCA AACACACCCC CTTAACCTAT CTCTTGGGAC TTAGCCCGAC	2040
	CCTCCCTCTC ATTTCCCATT AAGTCTGAGA GGCAAGAGCT AGGTTAGGCA AGGAGGTGGT	2100
10	TGGCCAGAGA TGGGGAACAG CCAGGTGCC CAGTCCTCTG ATTTTCCTC CATCCTGCTT	2160
	ACCACTCCC TGGGTACTTA CAGCCTCTC TGGGAACAG CGGGGGCCAG GACTGGGTCA	2220
	CCTATGAGCT GAATCAGCAT CTCCCTCTGA GTCCCAGGGC CCCTGCAGTT CCCAGTCTCT	2280
15	TCTGTCCTGC AGCCCTTGCC TCTTTCCCAC AGGTTCCACT TTATATCCAC CTTTCCTT	2340
	TGTTCAATTT TTATTTTAT TTTTTTATT ATTAAATGAT GTGGTCTATG GAAAAAAA	2400
20	TAAAAATCTG ACTTAGTTT A	2421

(2) INFORMATION FOR SEQ ID NO: 47:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 840 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

35	CTCAAACCTCC TGAGCTGAAG CGATCTACCT GCCTCAGCTA GGATTACAGG TGTGAGCCAC	50
	CGCACCCAAC CTCAATAAGC KTATTTGATA AAAKATATGC AAGCTCCCTT TATKCACCTT	120
	TCATTTCAGAA TGTTTAGTAA TTTGTATTGT TTTTCAGATT TTCAGCCAA TATATCTCCY	180
40	TGCCCACTGT GTCACGTAT TCTACCTAWA CATCATCACG TGTTCTGCT ATTGGCTGTA	240
	TGATGGAACA CTGGGGCTCA TTTTCCTGAA AACTGCCGAT AGTGCATAGA RTGCTGGGAT	300
	GGAAACCAGA ARCTTTGAAT TCAAGCCTTG GTCCTGCCPT GTTTTGCTT GGGTGGCCTT	360
45	GAGTCAGCCA CATACTTTT AAAATCTCAA TTTATTAGAA ATTATTCAA ATCAAAATCA	420
	AATGAGAAGG TATATACAAA AGTGTCTTAT CCCACAATAA ACTATTCAAG AGAGAGCAAA	480
50	GGAGAGGACA TTTACTCAAC ACCTCCTAAA AGGCAGCCAG TGAAATTAGG CATTTCATT	540
	AATCCTCCTG GCAACTCTGA GAGTAAAGCA TTATTAATCC CATTGGCT GTTTAAAGAA	600
	ATTATTTGCA CTAGATTCCA GCTGTAGTTT AGYTTCAAGA AAAAAAATCC TGAGATGTGA	660
55	ATTACACAGCT TTCTGGGTT AAAGCCCAAG CTCTATCACA TCATGCTATT ATTGTTACAT	720
	TACTGCTAGT TCTATGAAA GAAATACTAA TTTATGAAAT ACATCTTATC CAAAAAAA	780
60	AAAAAAAAAC TGGGAGGGGG GGCCCCGTACC CAAATCGCCG GATAGTGTAC GTAAACAATC	840

5 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2432 base pairs
 (B) TYPE: nucleic acid
 10 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15	GGCACGAGGC CGGAAACGCT GAGGAAGGGC CCGTCCCGCC TTCCCCGGCG CGCCATGGAG	60
	CCCCGGGGGG TTGCAGAACG CGTGGAGACG GGTGAGGAGG ATGTGATTAT GGAAGCTCTG	120
20	CGGTCATACA ACCAGGAGCA CTCCCAGAGC TTCACGTTTG ATGATGCCA ACAGGAGGAC	180
	CGGAAGAGAC TGGCGGASTG CTGGCTCCG TCCTGGAACA GGGCTTGCCA CCCTCCACC	240
	GTGTCATCTG GCTGCAGAGT GTCCGAATCC TGTCGGGGGA CCGCAACTGC CTGGACCCGT	300
25	TCACCAGCCG CCAGAGCCTG CAGGCAYTAG CCTGYTATGY TGACATCTCT GTCTCTGAGG	360
	GGTCCGTCCC AGAGTCCGCA GACATGGATG TTGTACTGGA GTCCCTCAAG TGCCTGTGCA	420
	ACCTCGTGCT CAGCAGCCCT GTGGCACAGA TGCTGGCAGC AGAGGCCCCC CTAGTGGTGA	480
30	AGCTCACAGA GCGTGTGGGG CTGTACCGTG AGAGGAGCTT CCCCCACGAT GTCCAGTTCT	540
	TTGACTTGCG GCTCCCTCTTC CTGCTAACCG CACTCCGCAC CGATGTGCGC CANAGCTGTT	600
35	TCAGGAGCTG AAAGGAGTGC GCCTGCTAAC TGACACACTG GAGCTGACGC TGGGGGTGAC	660
	TCCCTGAAGGG AACCCCCCAC CCACGCTCCT TCCTTCCCAA GAGACTGAGC GGGCCATGGA	720
	GATCCTCAAA GTGCTCTTCA ACATCACCCCT GGACTCCATC AAGGGGGAGG TGGACGAGGA	780
40	AGACGCTGCC CTTTACCGAC ACCTGGGGAC CCTTCTCCGG CACTGTGTGA TGATCGCTAC	840
	TGCTGGAGAC CGCACAGAGG AGTTCCACGG CCACGCAGTA ASCCTCTGG GGAACCTGCC	900
45	CCTCAAGTGT CTGGATGTTC TCCTCACCCCT GGAGCCACAT GGAGACTCCA CGGAGTTCAT	960
	GGGAGTGAAT ATGGATGTGA TTCTGCCCCCT CCTCATCTTC CTAGAGAACG GTTTGCACAA	1020
	GACACACAGG CTGAAGGAGA GTGTAGCTCC CGTGCTGAGC GTGCTGACTG AATGTGCCCG	1080
50	GATGCACCGC CCAGCCAGGA AGTTCTGAA GGCCCACGGTGTG CTGCCCCCTC TGGGGATGT	1140
	GAGGACACGG CCTGAGGTG GGGAGATGCT GCGGAACAAG CTTGTCCGCC TCATGACACA	1200
55	CCTGGACACA GATGTGAAGA GGGTGGCTGC CGAGTTCTTG TTGTCTGT GCTCTGAGAG	1260
	TGTGCCCCGA TTCTCATCAAGT ACACAGGCTA TGGGAATGCT GCTGGCCTTC TGGCTGCCAG	1320
60	GGGCCTCATG GCAGGAGGCG GCCCGAGGGC AGTACTCAGA GGATGAGGAC ACAGACACAG	1380

	ATGAGTACAA GGAAGCCAAA GCCAGCATAA ACCCTGTGAC CGGGAGGGTG GAGGAGAAC	1440
	CGCCTAACCC TATGGAGGGC ATGACAGAGG AGCAGAAGGA GCACGAGGCC ATGAAGCTGG	1500
5	TGACCATGTT TGACAAGCTC TCCAGGAACA GAGTCATCCA GCCAATGGGG ATGAGTCCCC	1560
	GGGGTCATCT TACGTCCCTG CAGGATGCCA TGTGCGAGAC TATGGAGCAG CAGCTCTCCT	1620
10	CGGACCTGAGCTGGACCCCT GACTGAGGAT GGCAGCTCTT CTGCTCCCCC ATCAGGACTG	1680
	GTCCTGCCTC CAGAGACTTC CTTGGGGTTG CAACCTGGGG AAGCCACATC CCACCTGGATC	1740
	CACACCGCC CCCACTCTC CATCTTAGAA ACCCCCTCTC TTGACTCCCG TTCTGPTCAT	1800
15	GATTTGCCTC TGGTCCAGTT TCTCATCTCT GGACTGCAAC GGTCTCTTG TGCTAGAACT	1860
	CAGGCTCAGC CTCGAATTCC ACAGACGAAG TACCTTCCTT TGTCTGCGCC AAGAGGAATG	1920
20	TGTTCAAG AGCTGCTGCCTG AGGGCAGGGC CTACCTGGGC ACACAGAAGA GCATATGGGA	1980
	GGCCAGGGGT TTGGGTGTGG GTGCACACAA AGCAAGCACC ATCTGGGATT CCCACACTGG	2040
	CAGAGCMANT GTKPTGGGGT ATGTGCTGCA CTTCCCAGGG AGAAAACCTG TCAGAACTTT	2100
25	CCATACGAGT ATATCAGAAC ACACCCCTCC AAGGTATGTA TGCTCTGTG TTCTGTCCT	2160
	GTCTTCACTG AGCCAGGGC TGGAGGCCTC TTAGACATTC TCCTTGGTCC TCGTTCAAGCT	2220
30	GCCCACGTGA GTATCCACAG TGCCCGAGTT CTGCTGGTT TTGGCAATTAA AACCTCCTTC	2280
	CTACTGGTTT AGACTACACT TACAACAAGG AAAATGCCCG TCGTGTGACC ATAGATTGAG	2340
	ATTTATACCA CATAACCACAC ATAGCCACAG AAACATCATC TTGAAATAAA GAAGAGTTT	2400
35	GGACAAAAAA AAAAAAAAAA AAAAAAAAAA AA	2432

40 (2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1742 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50	GTCCTGCAGG AGCTGCACCG GGCAGGAGGTG CGCANGAACAA AGGAGCAGCG AGAACAGATG	60
	TCGGGCTAAG GGCCCGGSAC GRGSGCGGCC CATCCTGCGA CGGAACACGT TCGGGTTTTG	120
	GTTTGTTCGTTTC GTTCACCTCT GTCTAGATGC AACTTTTGTT CCTCCTCCCC CACCCCAAGCC	180
55	CCCAGCTTCA TGCCTCTCTT CGCGACTCAG CGGCCCTGCC CTGTCCTCGT GGTGAGTCGC	240
	TGACCAACGGC TTCCCCCTGCA GGAGCCGCCG GGCGTGRAGA CGCGGTCCCT CGGTGCAGAC	300
60	ACCAGGCCGG CGCGGGCTGG GTCCCCCGGG GGCCCTGTGA GAGAGGTGGY CGTGACCGTG	360

	GTAAACCCAG GGCGGTGGCG TGGGATCRCG GGTCTTACG CTGGGCTGTC TGGTCAGCAC	420
5	GTGCAGGTCA GGGCAGGTCC TCTGAGCCCC CGCCCCCTGGC CAGCAGCGA GGCTACAGTA	480
	CCTGCTGTCT TTCCAGGGGG AAGGGGCTCC CCATGAGGRA GGGGCGACGG GGGAGGGGG	540
	TGATGGTGCC TGGGAAGCCT GCKTGTGCAN CCGGTGCTTG TTGAACTGGC AGGCCGGTGG	600
10	GTGGGGCTG CAGCTTCCT TAATGTGGTT GCACAGGGGT CCTCTRAGAC CACCTGGCGT	660
	GAGGTGGACA CCCTGGGCCT TCCTGGAAGC CTGCAGTTGG GGGCCTGCC TGAGTCTGCT	720
15	GGGGAGTGGG CATTCTCTGC CAGGGACCCA TGAGCAGGCT GCATGGCTA GAGGTTGTGG	780
	GCAGCATGGA CAGTCCCCCA CTCAGAAGTG CAAGAGTTCC AAAGAGCCTC TGGCCCAGGC	840
	CCCTCCGTGG GACAGCCCCG CGGCCCTCC CCACCAAGGC TTTCAGATG TCCTTGAAAG	900
20	ACCCACCCCTA GAGCCCTTTC GAGTGCTGGC CCCTCTGTG CCCTCTGCC TGGTGGAAAGC	960
	GGCASCACAA GTCCCTCTCA GGGAGCCCCA AGGGGGATTT TKTGGGACCG CTGCCACAG	1020
	ATCCAGGTGT TGGAAAGGGCA GCGGGTAAGG TTCCCAAGCC AGCCCCAACCA CCCCTCCCAC	1080
25	TTGGCACCCA GAGGGGGCTG TGGTGGAGG CCTGACTCCA GGCCCTCTCCT GCCCACACCC	1140
	TCTGGGCTGA GTTCCCTCTT TCCCTGGAC GCCCAGTGCT GGCCCTGGAG GACGGTCAGC	1200
30	TGGAGGATGG CGGTGGGGGA GGCTGTCTTT GTACCACTGC AGCATCCCCC ACTTCTCCAC	1260
	GGAAGCCCCA TCCCAAAGCT GCTGCCTGGC CCCTTGCTGT AAAGTGTGAA GGGGGCGGCT	1320
	GAGTTCTCTT AGGACCCAGA GCCAGGGCCC TCAACTTCCA TCCGTGGGA GGCCCTGGCC	1380
35	GGGCACTGCC AGTGTCTTCC AGAGCCACAC CCAGGGACCA CGGGAGGATC CTGACCCCTG	1440
	CAGGGCTCAG GGGTCAGCAG GGACCCACTG CCCCATCTCC CTCTCCCCAC CAAGACAGCC	1500
40	CCAGAAGGAG CAGCCAGCTG GGATGGGAAC CCAAGGCTGT CCACATCTGG CTTTTGTGGG	1560
	ACTCAGAAAG GGAAGCAGAA CTGAGGGCTG GGATATTCCCT CATGGTGGCA GCGCTCATAG	1620
	CGAAAGCCTA CTGTAATATG CACCCATCTC ATCCACGTAG TAAACTGAAC TAAAAAATTC	1680
45	AATCAAATGA ACAATTAAT AACACCTGT GTGTTAAGA AAAAAAAAAA AAAAAAAACTG	1740
	CG	1742
50		

(2) INFORMATION FOR SEQ ID NO: 50:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1487 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

	GGCACGAGCC TCCGCGAACT GTGGAGTCGG CGGAGGGCTG GAATCAGCGT GGGCTCCAGG	60
5	TCGCTGGCAG CCGGGTGGCA GAACTCTTCC GAGGCTCCTT GGGAAAGAAGC TACACCCGAG	120
	GGAGCCGGAT GGGCCTCGAA AACCTGGCCC GCTCTGGTC TGTACCATTG CAAGGGGAAC	180
10	CGTAAACTGA GCTTTCTAA CGTGGGTTTC TGCCAAGTAC TTTTCCAGCT GCCCCCCCTCC	240
	CCCCACGACA CAGGAGAGCC TCTGIGTAGC CAGCGCTTGA CAGTCGTTAG GTAGGTTGTA	300
	CTGTGTAGGG AGGAGCTCAA GATCATGAAT GGTTGTCACA GGAGAAAGCG GTTGCATCTT	360
15	TGCAAAACTA TATACTTGCT GTGGTTTGTG TTTTCTTTTC TGCTGAGTAA TGAAGTTGTA	420
	AGTTCACACT GGCACATTCT CAGGGCTGTG CAGATTATTT GCACTTTATT TCATAGGTGR	480
20	ATAAGTGCTT TTTAGCTTTC TTTGTATATT GAGTTGCTTT TGAATTGCTT CCCATATTTT	540
	TATTTCATAC AAACGTAAACA ATTGTGGCCC CTCTATTTTA TTTATAAAGG TTCACTGTAT	600
	CTTTGCCTGC CTACATCAAT CTGCAAGGGA GTGCCAGAAA GCCTCATGTT CATCGAGCCG	660
25	TGAGTCACAA CCAATTCTA AGCTGTTATA ACAAAAAAGT GTTTGCTTTT TTTCAACAAGT	720
	AACTTTAAAA GTGTAGTTTA GAAAGAAAAC ATTTCAATA AAAAGACACT ACATTAATCC	780
30	TGGATGCTTG CAAATCCTAA AATMTATTCC TCCCTCTAGCG TTGCACAGCT CTGTGTGTA	840
	TACACAGACT ACCTTTAAAA TTTGTCACAT ACCACTTTAC CTTTACTTTT ATGTATCATT	900
	CCCCCGACTT CCTTACTGCA GGTGTGGCA AGAAAACTT TCCTTTAACCA CTTTCAACA	960
35	GCGGGCATAA AATTCTGAG CTGAGGTCTT GAAGAATGCA GATGGGTACA GTATGTGTTG	1020
	GAGCTCACAG TGTTGTTGTA CTAACCTAGT CCCTTTTTTG CTTTTTTGG TATTGTCTTG	1080
40	TTAAAAGTGA CTCCCAGGTA GCAACTCTCT TTTTTAAGGG TGGGAACGAA AGGGACGTAG	1140
	GAAGAATAGA TCTAGATTAT TTAACAGTCT TCGATAGAGT TTGAAAGCTT TCTTCTTCAT	1200
	TCAATTGTTGG CCAAAATACT GCCTCTGCAT TTGTTACATA CAAAAGATT AGATTAATAA	1260
45	GTAGCTTTTG TTGGTGGAAA TTACCAAGCTC TATAAGTCAC CCTTGGTGGT TCATGGACCT	1320
	CTGATTAGCT TGGGTTTGC AGTCTCATTG CCACATGTAT ATGTGGAGCC AATGGCCTTT	1380
50	TGGTGCTCAG CTGTTTACGT CTGACTCCCTT GACTTCTTTG GTACAGTGAT GGAGTCAGAT	1440
	CTCATTAAGT GTGATTCTCC ATGGATATAA CCAGCCCCAA AAAAANG	1487

55

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1328 base pairs
- (B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5	GGCACGAGCT CGTCCCGAAT TCGGCACGAG AGAACATTG AAGAACCCAG ATCCAGCTTC	60
10	CCTGCGGGCT GCTTCTTGTG GGGAAAGGAA AAAGAGGAAG GCCTGTAAGA ACTGCACCTG	120
15	TGGCCTGCC GAAGAACTGG AAAAAGAGAA GTCAAGGGAA CAGATGAGCT CCCAACCCAA	180
20	GTCAGCTTGT GGAAACTGCT ACCTGGCGA TGCCTCCGC TGTGCCAGCT GCCCCTACCT	240
25	TGGGATGCCA GCCTCAAAC CTGGGGAAAA GGTGCTCTG AGTGATAGCA ATCTTCATGA	300
30	TGCCTAGGAG GTTCCCTGACA TGGGACCCAT CTGCTCCTCC AGCCAACCTCC TGCCCTCAC	360
35	ATCCCACCAT GGTCGGCTCCT CCCACCTCCT CTGGATTGT TCACTCTGAG ATCTGTTGC	420
40	AGAGTGGGTG CTTAGCAGAC AGAGTGAAGC TGGCTGGGG GCACAGTGGT GTGTAGTGCT	480
45	GCTGTGTATC AAAAGACCAA GGTATTATGG GACCTGGTTT CAGAATGGGA TGGTTTCTT	540
50	CACCTCATGT TAAGAGAAGG GAGTGTGTCC TGAAGAAGCC CTTCTCTGA TGTAAAATG	600
55	CTGACCAGAA CGCTCTTGAG CCCAGGCATC GTTGAGCATT AACACTCTGT GACAGAGCTG	660
60	CAGACCCCTG CCTTGAGTCT CATCTCAGCA ATGCTGCCAC CCTCTGTCT TTCAAGAGTTG	720
65	TTAGTTTACT CCATTCTTG TGACACGGT CAAGTGGCTC ACAACCTCCT CAGGGCACCA	780
70	GAGGAACACAC TCACTGGTTG CTGTGATGAT ATCCAGTGTG CTCCTGGCCCC CTTCCATCCC	840
75	CAACCACATT TGACTGTAGC ATTGCATCTG TGTCTGTGTC TCATTTATGT TAACCTTCAG	900
80	GTATTAAACT TGCTGCATAT CTTGACATAT CTTGAGATTC TGCAATGCTT GTAAAGAGAG	960
85	GGGATGTGCA TTTGTGTGTT ATGTTGGATA GTCATCCACG CTCAGTTGG ACCATTGGAG	1020
90	GAACCTAGTG TCACGCACAA ATGGGCTAT TCCTACGCTT AGAACATGGC TTGTCTGCC	1080
95	ACTTTAGAAG AGTCCCAGGT TGGTGAGCAT TTAGAGGGAA GCAGGGCAGA ACTCTGAACG	1140
100	ACAATACGTC TCTCTGAGCA GAGACCCCTT TGTCTTGTGTT ATCCACCCAT ATGGACTTGG	1200
105	AATCAATCTT GCCAAATATT TGGAGAGATT GTGTGGATT AAGAGACCTG GATTTTTATA	1260
110	TTTACCAAGT AAATAAAAGT TITCATTGAT ATCTGTCCTT GAAAAAAA AAAA	1320
115	AAACTCGA	1328

55 (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1856 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	GAATCGGCA CGAGCTCTGC AACATTGCCA ATGAACTTGT AGCCGAGGGT TCCGCTGCC	60
	CCTAGATTA ATTCCCCGGG CTGAACTGA GTGGAGATT TACATATCA TATTTTAAT	120
10	TGCTGTCTTC AATTACCCA TTCTATGCCA TAACTTAAT TCAAGGTGTC GATGCATGCT	180
	TTTCAGGCC TICCTTCITT GTACAAAGT AAATGTCAT AAACGTTTC ACTTATAATC	240
15	TTCAAACATG ATGCTTAATT AAATTAATA CTCCTATGC TACGTATTA TTCTATGAT	300
	TTTGCCACTG TTATTAAGTC TCTCAAAAT ACATCTAGG AACGGATTAA TTAAAGTRA	360
	TTTGATTATC TTCTATCTC TTCTTATTAA TTCTGATTC CTAGAAAT TCCTCCATT	420
20	GGTTGGCATT GATACAGTAA ATTTGTAAAT GAGGGAGACA TAAATAAAT CTAATTACT	480
	TGTGCTTAAT GACTGTAGCA GAATSCCTT TCTCTAAAT AGTTGCTT TCTTGCAGTT	540
	TAGTTTGATA GATTTGCAAG CTATGCTCT TCCATGAAAT TACGTGGCT CGTAGGAACG	600
25	CAGGCTTCIT TGTCTCTGGT TGTAAGCTTG ATGATGCCCT CTGGGCGAG ACAACGTAGC	660
	CGGAGATCAC AAATCAGGCC CTGGGTGAG TTGCTATGT CTGGAGGTGC AGAGAGGTTG	720
30	GCAGAAACTG ACCTCAGTGG GCAGGGTGG CCATGGACCT GATCTTTAA TGCACTCTAT	780
	GTGTTAGGA AGCCACGGC CATATTCAC TCTGAGAAAG AACAGAGG GAAAAACCCC	840
	ACAAAGTATA ACACCCCTT AAGATACATC TATTTAAAC TGAATTAAT TTTTCAGTTT	900
35	ATACCATTGG CCATTACAA GATAGGATG TTCATTCTC TTAAGATCC TTTGTTGACT	960
	TGTCTTTCA TCTCTTGCTA TTATTAATG TCACTGTTG TCAAGAAAGT CTTATTTGCT	1020
	GAGGAAGGAC TTTGCTGCAC TTACTGACCC ACATCAAGC CTGGAGGG TGGTGTAAA	1080
40	CTTTTTAAA AATGTTATTC TGATTAAAC AAATTAATG GCTTTTCA TGAAAAGAGC	1140
	GCCACCTTGC AAGGTTAGT GAGATTCAG GATCTGAA ACCTAGCAG GAATTGCTGC	1200
45	TAGCTCCAAA AATTGCGAA GCPAAAGCTA GCCCCATTCG GTTGGAAAGT TTGAAACTGA	1260
	TTAACAGATT TGCATTGAA GTGACTCGG ACATTAAGGT CCGATTTG TAAAGAATAG	1320
	AAAGAGGAAT AAAGACATCT YTTCCTCTCA GAAAGTAAAC CACCAATT AATATCCTT	1380
50	CCCACTTCA TTGAGATCAG CTTGCTCTAT AACCTGATC GAGGTGACA ATGATAAAC	1440
	TGATAATAGT CGTACTTTTG TAATTTGCT CGTGCATTG AGAGATAGT AAAGGATGAG	1500
55	TTCAYCTTT CTYOGAACAT YCCTATCCT AGATGCTAGT TACCTCAATT TGGGAATTAT	1560
	AACTGTCTTA ATTTTGTTG TGTACCTGGA TGCCCTTTT GCTTAATTC CCACAGTGTA	1620
60	ACAATTAAAT ATCACCTAT GACATAGTGT TTATGAGG TTTTAAGG ATAAATTATA	1680

GGGGTAAATG TTTACTTCAA AATGACTCCA TATTCAAAT ATCTGTITAG ACTGTGAAGG	1740
CCAATAATT TTTAAGAAAA CATTGAAGA GTAGTGTGTT TGCATTGTG AATAATCTTA	1800
5 CTCACAGCAA GTAAACGTAA TAAAAGCCAA CATTAAAGCC AAAA AAAAAA AAAAAA	1856

10 (2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1558 base pairs
- (B) TYPE: nucleic acid
- 15 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

20 TGGGTATCCA TTCCTGNAAT TACTTTACTT AGGATAATGG CCTCCAGCTC CGTCCAAGTT	60
GCTGCAAAAG GTATTATTC GTCCTTTTT GTGGCTGAGT AGTATTCCAT CGTGTATATA	120
TACCACATTT TCCTTATCCA CTCAATTGCTT GATGGCAGT TAGGTTGGTT CCACATCTTT	180
25 GCAATTGTGA GTTGTGCTGC TCCAGATATC ATCTTTAACCT CCTTTGCCCTT CTCCACATAC	240
ATTTCCAAGT CCTGTTCAATT CTACCTCCAA AATGTATCTT GTATCCATTG ATCTCTCTCC	300
30 ATCTTCAATC TATTTCAATG CCCCATCATC TCTTGATGG AGGAGTGTAA TAATTGGCTA	360
ACTGGCCTGT TCTTACATT TAAAATCAAA AGATGTGACA GGTGAAATGC CTATTTCACT	420
GTCCATTGAT GGTTCTGCTT ACACACCACC TGGCTGCCTG GTGTCCAGT GGCAGAGTTG	480
35 AGCAGTGTGA AAAAGACTGC TTGGCCCTTT ACAGGGAAAG CAGGTCCACT GTGGCCTGTG	540
AGGACGAGAG CTCTGGCAG GCTCGGACAC TGGCAGACCC TGGTCTGGC TG GCCAAGGC	600
40 AGCAGGGTAT GTGTTTGGG TCACTCACAG GGCTCAGCAC CACTCCTCAT GGCTTCCCTTA	660
CTGTTTCGGC AGAGGCTGAC CGOGGGCTGA TTGAGTCCCT CTCCCAGATG CTGTCCATGG	720
GCTTCTCTGA TGAAGGCGGC TGGCTCACCA GGCTCCTGCA GACCAAGAAC TATGACATCG	780
45 GAGCGGCTCT GGACACCAC CAGTATTCAA AGCATCCCCC GCGGTGTGA CCACCTTTGC	840
CCACCTCTTC TGGTGGCCCC TCTTCTGTCT CATAGTTGTG TTAAGCTGTC GTAGAATTGC	900
50 AGGTCTCTGT ACGGGCCAGT TTCTCTGCCT TCTTCCAGGA TCAGGGTTA GGGTGCAAGA	960
AGCCATTTAG GGCAGCAAAA CAAGTGACAT GAAGGGAGGG TCCCTGTGIG TGTGTGTGCT	1020
55 GATGTTTCCCT GGGTGCCTTG GCTCCTTGCA GCAGGGCTGG GCCTGCGAGA CCCAAGGCTC	1080
ACTGCAGCGC GCTCCTGACC CCTCCCTGCA GGGGCTACGT TAGCAGGCCA GCACATAGCT	1140
TGCCTAATGG CTTCACATT CTCTTTGTGTT TAAATGACT CATAGGTCCC TGACATTTAG	1200
60 TTGATTATTT TCTGCTACAG ACCTGGTACA CTCTGATTTT AGATAAAGTA AGCCTAGGTG	1260

TTGTCAGCAG	GCAGGGCTGGG	GAGGCCAGTG	TTGTGGGCTT	CCTGCTGGGA	CTGAGAAGGC	1320	
5	TCACGAAGGG	CATCCGCAAT	GTTGGTTCA	CTGAGAGCTG	CCTCCCTGGTC	TCTTCACCAC	1380
	TGTAGTTCTC	TCATTTCATA	ACCATCAGCT	GCTTTTAAAAA	TAAGATCTCT	TTGTAGCCAT	1440
	CCTGTTAAAT	TTGTAAACAA	TCTAATTAAA	TGGCATCAGC	ACTTTAACCA	AAAAAAAAAA	1500
10	Aoooooooooooo	AAANAAAAAA	AAAAGGGGGC	CGCTCTAGAG	GTCCAAGTTA	NGACGNNG	1558

15 (2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 948 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

25	TAAAAATCAT	GCTCTGTACC	ATCCTCACCG	TAGTCATCAT	CATGCCCGCG	CAGACCACGA	60
	GAACTACTGG	GATCCCTAAA	AACGCCCTTG	GTCCGGCCCC	ACTCTGCGCC	CCTCGATCTC	120
30	CCAGGCTCTT	TCTGCAGWCA	TACCGCGGAC	CCAATGGGCG	CCCTGCACAC	CCGTTTCTGG	180
	GGCCGTCAGA	CTTGGATACA	TCGTAAACTC	CCGCTCCACG	GAACGTCTCG	CCTKGCGAGC	240
	AAGMTGGAA	TCCAGTTCT	CAGGAACCCC	TCCAAAACCC	ACACCCCCAG	GGACGCCGCT	300
35	TTCCGGGATC	CCGGSCAAC	GCCGGACCC	CAGTCGCTCC	AGGCCCCCTC	ACCCCTCAAAG	360
	TGTAGCGCCC	CCAACCGAGC	AACCTCGGTT	TGGTCCCTAA	AACCCCGCT	CCTCTATAAG	420
40	CACCGCCCCA	GCTCTGACAA	AACCCCGCCT	CCAGGTGGC	AGGCTCCGCT	TCTTTCTTC	480
	TCCGGGGGT	GATTCAAGTCC	AGTGATIGGG	TTTGTGGCTC	CAGGCCTCGC	CCACAGACGG	540
	ACAGACCCCT	CCCTTTCTTC	CGGCAAAAGG	ACCGAGCCCT	GGGGTAGTAA	GGSCCCCCACA	600
45	CTCCTGTTTT	TTGCAAGTAC	ATTTTGTCC	YTCCTCCACC	CAGGTATCTG	CCTATTTCT	660
	TGCTAATCCC	AGAACCTTTC	CTTTTGCTTT	TTTTAAGGAC	ATTTGGAAG	TTCTGGTGT	720
50	AGGACCCCTTC	TCCCTGGGAT	AAGAAACCTG	CCTGTAAACG	CTCTGTAAAT	ACTCCCTTCC	780
	ACCCATCCCA	GCCCCCTGGGC	AGCCGGGCAG	AAGGGAATCC	AGGCTATGGA	CCTCCCAAGT	840
	CCCCGCTCCC	CGCTCCCCCTC	GGGGGCCCCG	CCTTGTTCTG	ATCTGTGTGT	GAGTGTGTGT	900
55	GAACTTCTGA	AAGACAATAT	TAAAGAGACT	TAGTTGAAAAA	AAAAAAAA		948

60 (2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10	GGGAACTGC AGTGACAGCA GGAGTAAGAG TGGGAGGCAG GACAGAGCTG GGACACAGGT	60
	ATGGAGAGGG GGTTCAAGCGA GCCTAGAGAG GGCAGACTAT CAGGGTGCAG CGGGTGAGAA	120
15	TCCAGGGAGA GGACCGGAAA CAGAAGAGGG CCAGAAGACC GGGGCACTTG TGGGTTGCAG	180
	AGCCCCTCAG CCATGTTGGG AGCCAAGCCA CACTGGCTAC CAGGTCCCCCT ACACAGTCCC	240
	GGGCTGCCCT TGGTCTGGT GCTTCTGGCC CTGGGGGCCG GGIGGGCCCA GGAGGGGTCA	300
20	GAGCCCGTCC TGCTGGAGGG GGAGTGCCCTG GTGGTCTGTG AGCCTGGCCG AGCTGCTGCA	360
	GGGGGGCCCG GGGGAGCAGC CCTGGGAGAG GCACCCCTG GGCGAGTGCC ATTGTYTGCG	420
25	GTCCGAAGCC ACCACCATGA GCCAGCAGGG GAAACCGGCA ATGGCACCAAG TGGGCCATC	480
	TACTTCGACC AGGTCTGGT GAACGAGGGC GGIGGGCTTG ACCGGGCCTC TGGCTCCCTC	540
	GTAGCCCCCTG TCCGGGGGT CTACAGCTTC CGGTTCCATG TGGTGAAGGT GTACAACCGC	600
30	CAAACGTGCC AGGTGAGCCT GATGCTGAAC ACGTGGCCTG TCATCTCAGC CTTTGCCAAT	660
	GATCCCTGACG TGACCCGGGA GGCAGCCACC AGCTCTGTGC TACTGCCCTT GGACCCCTGGG	720
35	GACCGAGTGT CTCTGCGCCT GCGTCGGGGG NAATCTACTG GGTGGTTGGA AATACTCAAG	780
	TTTCTCTGGC TTCTCATCT TCCCTCTCTG AAGGACCCAA GTCTTCAAG CACAAGAAC	840
	CAGCCCCCTGA CAACTTCTT CTGCCCTCTC TTGCCCCANA AACAGCANAA GCAGGANANA	900
40	NACTCCCTCT GGCTCCTATC CCACCTCTTT GCATGGGAAC CTGTGCCAAA CACCCAAAGTT	960
	TAAGAAAAAA ATAAAACGTG GGCATCTCCA	990

45

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1603 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

	GGTCGACCCA CGCGTCGGC CGCGCGGCTC CGGACGGCT CTGCCTICCC GACCGCGGGGA	60
55	CCCGCGCCCTG GGGGAGGAGG GCGAACGACG CGGCGATGGC TCCGCGGGCA CTCCCGGGGT	120
60		

	CGCCCGTCCT AGCCGCTGCT GTCTTCGTGG GAGGCCGCGT GAGTTGCCCG CTGGTGGCTC	180
	CGGACAATGG GAGCAGCCGC ACATTGCACT CCAGAACAGA GACGACCCCG TCGCCCCAGCA	240
5	ACGATACTGG GAATGGACAC CCAGAACATA TTGCATACGC GCTTGTCCCT GTGTTCTTTA	300
	TCATGGGTCT CTTTGGCGTC CTCATTNGC CAMCTNGCTT NAAGAACAGA GGCTATCGTT	360
	GTACAAACAGA ACCAGAGCAA GATATCGAAG AAGAAAAAGG TTGAAAAGWT AGRATTGAAT	420
10	GACAGTGTGA ATGAAAACAG TGACACTGTT GGGCAAATCG TCCACTACAT CATGAAAAT	480
	GAAGCGAATG CTGATGTYTT AAAGGCATG GTAGCAGATA ACAGCCTGTA TGATCCTGAA	540
15	AGCCCCGTGA CCCCCAGCAC ACCAGGGAGC CGGCCAGTGA GTCCTGGCT TTGTCACCAG	600
	GGGGGACGCC AGGGAAAGCAC GTCTGTGGCC ATCATCTGCA TACGGTGGGC GGTGTWGTG	660
	AGAGGGATGT GTGTCATCGG TGTAGGCACA AGCGGTGGCA CTTTATAAAG CCCACTAAC	720
20	AGTCCAGAGA GAGCAGACCA CGGCGCCAAG GCGAGGTCAC GGTCCCTTCT GTTGGCAGAT	780
	TTAGAGTNAC AAAAGTGGAG CACAAGTCAA ACCAGAACGA ACGGAGAACCC CTGATGTCTG	840
25	TTAGTGGGCC TGAAACCGTC AATGGGGAGG TGCCGGCAAC ACCTGTGAAG AGAGAACGCA	900
	GTGGCACAGA GTAGCAGGTG AGCCGTGGTT TTGGTGACAT TGGGGCAGA GTGGTGCAGG	960
	GTGAGGAGAA GGTACTTGGGA GCCTCCCAGG TGCTGTGGCA GCATAGGAAT GGTATTTGAC	1020
30	AGGGAAAGTGG GAGAGCTTTC CTTGACCCAG GAAGACTGAG GGGGACTGAA CATGATTACT	1080
	TGTCTGCCTA GAGCTTCTTG TAAAGAACGTC ACAAAACTTAG TGCCTCCAGG GGCTTGGCTG	1140
35	TGTGATAATG AGGATAGAGG ATTACTGTG AGGCAATGTG GCATGGTGGG GATTGTGGCA	1200
	AACTAGAATT CACATCACCC ACCATATAGG GCTTGCATTA CCACGAGGCA GAAAGCACCT	1260
	AGTGTGCTG CATCTCTTA CGCAAAAAAG ACAAAATCCA GACTTCTAAA ATGTAAAATC	1320
40	ACTGATTTTC GATATTGGCA GCTTACTTTT TTTTTTTAAA CAACCATGCA GGCCAAATGA	1380
	CTTGTAATCT TGTCAACCATT TTTAGGTAAA CTGTGACTTG AAAAAGTCTG GAGCAAACAA	1440
45	ACCAATGCTT TTTCCTTTTA TTCTGTGGR AACCAAGTTT CTTTGTGTCA CAGTTYTGAA	1500
	ACCTCAATAAC GAATATTTCT CTTCCCACCA AATATTTGA GGCAATTGAA AAGCCACAGT	1560
	GATTTATTTTC TTGATTTGGC AATTTAATT TTGCAAGACA ATT	1603
50		

(2) INFORMATION FOR SEQ ID NO: 57:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1052 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

	TACAGCTCAG GATGCCGTGA ACATTTGTCAT CTCTGGGCTT CTGGGTCCCTG CTTAGCCTGC	60
5	TTTTTCCCTG GAGGACTGAC CAGGGATGCG GCCCAGCAAC ATGTTACTAA ATCATACTCT	120
	CCTCCCTACC TTTCCCAGAC CTCTCACTCC TGCCCTGGTGT TCCAACCCGT TCTGTGGCCA	180
10	GAGTATAACAT TTTGGAACCT CTTGAGGCC ATCCTGCAGT TCCAGATGAA CCATAGCGTG	240
	CTTCAGCAGN AAGGCCCGAG ACATGTATGC AGAGGAGCCG AAGAGGCAGC ACCTGGAGAG	300
15	GGACCAGGCT ACAGTGACAG AGCAGCTGCT GCGAGAGGGG CTCCAAGCCA GTGGGGACGC	360
	CCAGCTCCGA AGGACACGCT TGCACAAACT CTCGGCCAGA CGGGAAGAGC GAGTCCAAGG	420
	CTTCCTGCAG GCCTTGGAAC TCAAGCGAGC TGACTGGCTG GCCCGTCTGG GCAC TGCCATC	480
20	AGCCTGAATG AGGCTGGCCA CCTGCCACTT TGCCCTGCC TCTGCCTCCA GGGCTCCMCT	540
	MYCCTTCCTT TTCTTGGTGA AAGGCACCTC CTTTCCCTGAT AATGAATGGT GTTCCCTTTC	600
	CTTGGCTGGG GAGCCCCCA GGCCAGGTTT GCTGGCCATA GATACCTTGT GGCTGCCTGR	660
25	GACAGGCTCC TGAGGAGGAT TGAGGGTGAA AGTCTCCCAC GAGTACACTA AACCTAGGTC	720
	TGGTCACCAA TAGGGTTTGG AGAGCAAAGG GCCACAACTC ATCAGCTGCC TGTCTCTTAG	780
30	ATGCACTTTC TTTTCCACC AGCACATCCT TCAACACACA GAATTCAGG GAAGAGTTCT	840
	CCCCAAAACC CTAGCTCTTT ACCCTTCCAT TTAGCCCTTC CACCCAGCTT CCACAAAAGA	900
	TTTGGCTCTA CCTTGGATCT GCTAGTAAAT AACTAATAGG CAGGCAGTTA TTTGGGTAAAG	960
35	GAAAAAAGGG GTGGGAGAGA CAGAAAATTG GCCCACTGCT GCTCCCTCCCC TTGGSTYTCC	1020
	ACCTGGGATT TGCTATTGAA TCTCTACCCCT NN	1052
40		

(2) INFORMATION FOR SEQ ID NO: 58:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 814 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xii) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

	ACNCGNIGGC GGCCGCTCTA GAACTAGGGG ANCCCCGGGG CTGCAGGAAT TCGGCACGAG	60
55	CATAGACTTT TAAACTGGTA CGGTTCTTAG AGATGGTCCT TGGCCTTCTG TTGTTGTTGT	120
	KGTTTTTTTC TTTTCTCTCT TCTCTCTCTC CTCTCTCTC TCTCTCTCTT CTCTCTCTT	180
	TTTTTTTTCA GAGTCTTGCT CTGTCACCAA GACTGGAGTG AAGTGTATGTG ATCTCGGCTT	240
60		

	ACTGCAACCT GGGAGGCAGA GGTTCCAGTG AGTCGAGATG GTGCCATTGC TCTCGTTTGG	300
	GCAACAAGAG TGAAACTCTT GTCTAAAAAA AAAAAAAAAA ATGAGGTTA AGACAGTTTT	360
5	GTCATTACTG GTGGGATCTG GTCACACAAAG ATAGCATTAA ACGTGACATG GCACATAAAA	420
	TTGGTTAAA AATTTGTTT TTTAATTACG TAATGTAAA GCCAACAAA CACTTATGC	480
10	AAGATTGGAA TGTATCTTCA AATTCAAGATT TAATAAACAT GTAAAGATCC TCTGTATATA	540
	AAAGTGTAT TTAATCCCTT GTCCCCAAG AATGCTATAA AAGATCCAA GAATGTTATC	600
	TATGAAAAGA TAGCAATAGG GAATGGTGA CAAATAATT AATTTGCCAA TTCTAAAAAA	660
15	CATGGACTTA AACCCCATGA AAACCTGGTT CCATAGTTTT AACTGTTTAA TGGTTCCAAT	720
	ACAAAACCAAG AGTGGTTAC ATTCCACAAT NACCAAATT GCATCCAATN TTGGGGTAAT	780
20	TTTNGGTATT TGCCATGGGA TACTATTCA TTTT	814

25 (2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1215 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

35	AGAGGAAGTC TTTGCCAAG CCTGTTCTCT GGACTAACGC CATCCAGGCT GGGAGGGAA	60
	GAGTGCTCTG CTACACTCGT CCCCCCTCCTG CCTCATCTTC CTTCTCAGCC TTGGTTCCCTG	120
	ATGGGAACAG AATGGAGGGC CTGAGAACAT ACTTTCTAAA TGCCCTTGAC CCAGGAACCG	180
40	ATTATCTATA TTGTTCCCA TTTCCCTCA CCGTGACATT CCAGCATTTG CTGACTGTGA	240
	GGTGGGCCCTTGAGAGCCTTC CAGGTTCCCTC AAAACAGGCC TGAGCGATGG GCATCACACC	300
45	CTCTGCCTAC CCACRTGCCT GCTTACCTGTC CAGATAACCA AGTGNAGATG TCTGCGAGTG	360
	GCTAGTTTTC ACATTCTTAC TAGTGTGTTGG YTACACCTTTCG GGCAAAGGCC CCCTCTAGGC	420
	CTTGCCCCAC CTCCATCAA CGCAGACACT GTAGTCAGAC CTCAGYAATA TAGGAGGCAA	480
50	TAATCTTTA ACAGTGTGTTT GCAAACAAAC AAAAAGAGAA AAATCCCAGC CAGGGAACT	540
	CGCCACCTGTC CCACGCTAGT TCCATCCACG CTCAAGACCC GCCCTTAGAC CAGGCAGGCA	600
55	AAGGCCCCCA TCACACTCGG CCACTAGTGG GGTCTTGAGG CCAAGAAAGA AACCAGACCC	660
	TGTATGACAA GTGGGKTCT TTCCAGAACCA CGACAGAAAC AGGGGGGCC CCTTGTAAAT	720
	GCCACTCCAT ACTCCAGAAG CATTATTCT TATTTGGAC AGCCAAGGGC AGATTACACAG	780
60	GTTATGTAG GAATAAAGAC TAGTTTACAA AGGARAAAGA GSCTCTGGAC TTCCCCMAGGA	840

	AAGGTCAGGT TAGGGCTCCT GTACCCATTC TGTTCCACCA CTGTTGATC TCTCTGGCCT	900
5	CCCACCAGGA ATGCCGTTTC CTTTTTATGG ATCTGTTGGG AACCAGAGAG AATCAACAGA	960
	TCAATGACAT AGGATCCGAA GTGCAATGAT AGTCACCTCT AGTTTGGCAT TTCACAAACT	1020
	CTGNACAGCA AGGTATTGGT AGGTTACTCA ATTICAAAAG GGCCCCATGG CCAAATATGT	1080
10	TTAGGAACCG CTGTTGNAT TTCTTTTTTG GGAGACGCAT TGTATATAAT ATATGTCAA	1140
	GGCTTTCGGA ATTCCCTGCAG GAAAGAAATC AGCTTTGTTA AATCCNAAAA AAAAAAAAAAA	1200
15	AAAAAAAAATAG ACTCG	1215

(2) INFORMATION FOR SEQ ID NO: 60:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 478 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

30	ATTCTTATG ACATGGGGT TTGAATTGGT TGGCAAATGT TTAATTITAA TATCCATAAT	60
	CAGTGAGGTC CTGCTGGCTG TAATCATTAA TTGTGAAATC TAAGGAGCTT AGTTCATGGC	120
	TCTAGAATTTC CACAGAAAAR TGYGMTATGA TACGAGCATT AAGTTTATTT CTTCTGATCT	180
35	TTGATGCAGC TTGTTCACT TTATCTGTT TTGTATTAT TGGTCATCTA CTTCCCATGC	240
	CAAAAGGGAC TGGTCTACAT AGCTGGCTA AACACCTGAT CAAATCACTA AAAGAAAATG	300
40	TGTTACCTCT AATGAATTAT CCTGATTGTA AGTTAAAAT CAATATTCC CGTAGTGAG	360
	GTTTGCTTTT TAAAAAGAAK KCTTAAAAAA AAAAAAAAAA AAACGAGTTA AAGAAAAGGA	420
45	AGCAAGCTCA GGTAAGGTGC ACACATTGGG CTAAGGAAGC TAGAGCCTGT GGAGANGC	478

(2) INFORMATION FOR SEQ ID NO: 61:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 618 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

60	TATGACCTTG ATAACCCCAA GTTNGAAATT AACCTTCANT AAAGGAAACA AAAGCTGGAG	60
	TTCGGCCCGCT TCCAGTTCGA CACTAGTGGG TCCCAAAGAA TTCCGGCACGA GTCATAATGA	120

320

GCTACTAGGT AAGCCTTCTG GGACTTTCAAG ATATTTGGG GAAGATTGAT TTTTGTTC	180
ACATGCTGTG GACCCTTGGC CATCAAATGG TATGGGAAG CTCATCCGTC TGTCTGTGAT	240
5 GGTCACTGTCA GTCAGGCGTC TTTTAGTAT TTACTGGGTG CTCAGTACTG TGCCAGATGC	300
TGTGGGAGC CGTGGTGGTA TGGAGGGAGGA GTGCTCCAGA GGACTCTGCT GTGTGGCAGG	360
10 CCAGCATAAA CAAGCCAAGG GGAAAAGGCA GGCATGGAAT AAACCCGGAG AATACCAGTG	420
TGTGACTTAC TGCTGACTGT GTGGATTAGC CTATCAGCAG TAATCAAGCA GGGCGGAGGG	480
15 CATTATCTTT GAGCCAGAAG AGTGAGCACT GSSCCGAGGG TGGACCATCA AGAGGGGGTG	540
TAGGACCNCA AGGCTTCTTN CNGGGGAGAC AACGTCAATA AGCNGTCAGT AGTCACCGAC	600
AGTTTGGGA AGCAAGGG	618

20

(2) INFORMATION FOR SEQ ID NO: 62:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 751 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TCGACCCACG CGTCCGAGGA GCTGGACITC TGAGACAGCC ATTCTCTTG CATAGCACTG	60
35 TCTGCTGCTA CAGCTCATAG AAGTCAACAA TTTCTTCAA CACTGGTAGG CAGCCTCTAA	120
ATGGCCCTGA TCACCCCTCAC CTCCCTGCCAT TCACACCNNT GTAAAATTCC ACCCCTGGAC	180
40 CTAGTGACTC ACTTCTAACAA ANGAGAATAC AGCAAAAGTA ACATGCCITC TGAGGTGAGG	240
CTACAACGGAG ACTACGATGC CTGCTTGGT CACCCCTCTC CTGCTCTTTC CATTGCTCCC	300
TCTGATGGAA GCCAGTTGCC ATGTGATGAG GTGCCCTATG GAGAGGGCCA CGTGACAAGG	360
45 TATTGTAAGAA AGCCTCTGAC CAATAGCCAT CTAGAAACGG AGGCCAGTC CAGCAGCCTC	420
TGAGATGAAT CCTGCCAACC TGAGCTTGGA GACAGATTCT CTCCCTATCC TGCCCTGGGA	480
50 TGATCACAGC CACCACCAAC ACCTTCACTG CCTGGTGAGA GGCAAGCCA GTGAACCCAA	540
GGTAAACTGG ACAGAACCT GACCCACAGA AACTGAGATA ATGTTGTTA TTTTAAGCTG	600
CTCAGTTGT TACAGAGCAA TAGATAACTA ACTCAAACAC CATAAAATTG TAATATTTA	660
55 TTCTATCACA CAAACCAGGT AATACCAAGT AAATGCCATT ACTATACACA TATTTTGTG	720
ACACAATTAC ATGTGATTG TTAAGAAGGC T	751

60

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 780 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CNGYAGTCA	CAGTCGGCGA	TTCGGGGTC	GACCCACGCG	TCCGGGTGG	CAACTCCTGA	60	
GGCCTGCATG	GGTGACTTCA	CATTTTCCTA	CCTCTCCTTC	TAATCTCTTC	TAGAGCACCT	120	
15	GCTATCCCCA	ACTTCTAGAC	CTGCTCCAAA	CTAGTGACTA	GGATAGAATT	TGATCCCCTA	180
	ACTCACTGTC	TGCCGTGCTC	ATTGCTGCTA	ACAGCATTGC	CTGTGCTCTC	CTCTCAGGGG	240
20	CAGCAGTCTA	ACGGGGGCGAC	GTCCTAACCTC	AACTGGGAGA	AGCCTCAGTG	GTGGAATTCC	300
	AGGGACTGTC	ACTGTCAAGC	TGGCAAGGGC	CAGGATGGGG	GGAATGGAGC	TGGGGCTTAG	360
25	CTGGCGGGTG	GTCTGAAGCA	GACACGGPAT	GGGAGAGGAG	GATGGGAAGT	AGACAGTGGC	420
	TGGTATGGCT	CTGAGGCTCC	CTGGGGCCTG	CTCAAGCTCC	TCCTGCTCCT	TGCTGTTTC	480
	TGTTGTTTG	GGGGCTTGCG	AGTCCTTGTG	TCCTCATCTG	AGACTGAAAT	GTGGGGATCC	540
30	ACGATGGCCT	TCCTTCCCTCT	TACCCCTCCT	CCCTCAGCCT	GCAACCTCTA	TCCCTGGAACC	600
	TGCTCTCCCT	TTCTCCCCAA	CTATGCAACT	GTGTCCTGCT	CCTCTGCAA	GGCCAGCCAG	660
35	CTTGGGAGCA	GCAAGAGAAAT	AAACAGCATT	TCTGATGCCA	AAAAAAAAAA	AAAAAAAACC	720
	GCGCCCGAAA	GCTTAAGTAA	GGGGTTAATT	TTTAGCTTGG	GCACTNGGCC		780

40

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 588 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TTCCGATTA	ATCGACTCAC	TATAGGAAT	GCGCTGCCA	TGACCCGCGG	TAACCAGCGT	60	
	GAGCTCGGCC	GCCAGAAGAA	TATGAAAAAG	CAGAGCGACT	CGGTTAAGGG	AAAGCGCCGA	120
55	GATGAGGGC	TTTCTGCTGC	CSCCGCAAG	CAGAGGGACT	CGGAGATCAT	GCAGCAGAAG	180
	CAGAATAAAGC	CAAACGAGAA	GAAGGAGGAA	CCCAAGTAGC	TTTGTGGCTT	CGTGTCCAAC	240
60	CCTCTTGGCC	TTCGCCTGTC	TGCCTGGAGC	CAGTCCCACC	ACGCTCGCGT	TTCCCTCCGT	300

AGTGCTCACA GGTCCCAGCA CCGATGGCAT TCCCTTGCC CTGAGTCTGC AGCGGGTCCC	360
TTTGTGCTT CCTTCCCCTC AGGTAGCCTC TCTCCCCCTG GGCCACTCCC GGGGGTGAGG	420
5 GGGTTACCCC TTCCCAGTGT TTTTTATTCC TGTGGGGCTC ACCCCAAAGT ATTAAAAGTA	480
GCTTGTAAAT TCCAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA	540
10 AAAAAAAAAA AAAAAAAAAA AAAANNCGGG GGGGGGCCCC CCCCCCCCC	588

10

(2) INFORMATION FOR SEQ ID NO: 65:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 774 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

25 TTTAAAGATG AAGAAATGAC AAGGGAGGGA GATGAGATGG AAAGGTGTTT GGAAGAGATA	60
AGGGGTCTRA GAAAGAAATT TAGGGCTCTG CATTCTAACCC ATAGGCATTG TCGGGACCGT	120
CCITATCCCA TTTAATTAAT TTCTCTGACA ATTCAATTAT TTTCTGTTAT TAATGTTGCC	180
30 ACTGCTTCT GTTGTCTGC ACTTTCTTGA TAAATATTG CTATCGTTT ACTCCAGTCA	240
TTCGATGTTG CTGAGATTTA CATATGACTC TTGTCAACAT CTCATCTTTT GACCCAATCT	300
TATTCATTTA ATAAGAGGTC TCATTCAATT GCATGGAAAA ATGCTCATTG TATATTGCAA	360
35 AGTGAAAATA ACGAGTTGCA AAACAGTGT A TACATATAAT TGTGTATATA TGTACACTTT	420
ATTGTACAT TTCTATGTGA CATAATGCAA AGGAAAGTGT CTGATTATAT TATACACCAA	480
40 AGGTAAACAG TGAATCTCTG TGTGATCTCT TTTTTTTCT TTTTGCCTAT CTGCATCTTC	540
TCACTTGCCA AAAATGAAT ATATGTTAT GTGTGTATAT TACTTGIGTC ACAAAAAACC	600
45 CTAAGTAGA CAGTAAAAGA ACTTGTCAAT CCCCTTGGA AGGCAATGAA ACACCTAATA	660
AACTCTCAAT AACAGAACCG TAAAATGAA ATGAAACCT CCAATTACCT CTGGATCTCT	720
TAGCCAGAGT AATAAACTGG TAATTATTAC AGATAAAAAA AAAAAAAA AANA	774

50

(2) INFORMATION FOR SEQ ID NO: 66:

55

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1866 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	ACCCACCGGT CGGGCCTCT TCTTCAGCAC ATGCCAAAGC TGTTCTCAC GGCTGTGAG	60
5	ACAAGAGCAT CTGGATGTA GGACAATGGA AGAGTTAGAT GCCTTATTGG AGGAACGTGGA	120
	ACGCTCCACC CTTCAAGGACA GTGATGAATA TTCCAACCCA GCTCCTCTTC CCCTGGATCA	180
10	GCATTCCAGA AAGGAGACTA ACCTTGATGA GACTTCGGAG ATCCCTTCTA TTCAGGATAA	240
	CACAAGTCCC TTGCCGGCGC ATCGTGTAT ACTACCAATA TCCAGGAGCT CAATGTCTAC	300
	AGTGAAGCCC AAGAGCCAAA GGAATCACCA CCACCTTCTA AAACGTCAGC AGCTGCTCAG	360
15	TTGGATGAGC TCATGGCTCA CCTGACTGAG ATGCAGGCCA AGGTTGCACT GAGAGCAGAT	420
	GCTGGCAAGA AGCACTTACC AGACAAGCAG GATCACAAGG CCTCCCTGGA CTCAATGCTT	480
20	GGGGTCTSG AGCAGGAATT GCAGGACCTT GGCAATTGCCA CAGTGCCCAA GGGCCATTGT	540
	GCATCCTGCC AGAAACCGAT TGCTGGGAAG GTGATCCATG CTCTAGGGCA ATCATGGCAT	600
	CCTGAGCATT TTGTCTGTAC TCATTGCAA GAAGAGATTG GCTCCAGTCC CTTCTTTGAG	660
25	CGGAGTGGCT TGGNCTACTG CCCAACGAC TACCACCAAC TTTTTCTCC ACGCTGTGCT	720
	TACTGCGCTG CTCCCACCTT GGATAAAAGTG CTGACAGCAA TGAACCAGAC CTGGCACCCCA	780
	GAGCACTTCT TCTGCTCTCA CTGOGGAGAG GTGTTGGTG CAGAAGGCTT TCATGAGAAG	840
30	GACAAGAAC CATAATTGCCG AAAGGATTTT TTAGCCATGT TCTCACCCAA GTGTGGTGGC	900
	TGCAATCGCC CAGTGTGGA AAACTACCTT TCAGCCATGG ACACTGTCTG GCACCCAGAG	960
35	TGCTTTGTTT GTGGGGACTG CTTCAACCAGT TTTCTACTG GCTCCTCTT TGAACGGAT	1020
	GGACGTCCAT TCTGTGAGCT CCATTACCAT CACCGCCGGG GAACGCTCTG CCATGGGTGT	1080
	GGGCAGCCCA TCACTGGCCG TTGTATCAGT GCCATGGGGT ACAAGTCTCA TCTTGAGCAC	1140
40	TTTGTGTGTC CTTCTGCCT GACACAGTTG TCGAAGGGCA TTTTCAGGGA GCAGAATGAC	1200
	AAGACCTATT GTCAACCTTG CTTCAATAAG CTCTTCCCCAC TGTAATGCCA ACTGATCCAT	1260
45	AGCCTCTCA GATTCTTAT AAAATTAAA CCAAGAGAGG AGAGGAAGG GTAAATTTC	1320
	TGTTACTGAC CTTCTGCTTA ATAGCTTAT AGAAAAAGGA AAGGTGATGA GCAAATAAAG	1380
	GAACCTCTAG ACTTTACATG ACTAGGCTGA TAATCTTATT TTTTAGGCTT CTATACAGTT	1440
50	AATTCTATAA ATTCTCTTTC TCCCTCTCTT CTCCAATCAA GCACCTGGAG TTAGATCTAG	1500
	GTCCCTCTAT CTCGTCCCTC TACAGATGTA TTTTCCACTT GCATAATTCA TGCCAAACACT	1560
55	GGTTTCTTA GGTTTCTCCA TTTTACCTC TAGTGATGGC CCTACTCATA TCTTCTCTAA	1620
	TTTGGTCTG ATACTTGTCTT CTTTCACGT TTTCCCATTT CCCTGTGGCT CACTGTCTTA	1680
60	CAATCACTGC TGTGGAATCA TGATACCACT TTTAGCTCTT TGCATCTTCC TTCAGTGTAT	1740

TTTGTGTTT CAAGAGGAAG TAGATTTAA CTGGACAACT TTGAGTACTG ACATCATTGA	1800
TAAATAAACT GGCTTGTGGT TTCAATAAAA AAAAAAAA AAAAAAAA AAAAAAAA	1860
5 AAAAAA	1866

10 (2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1152 base pairs
- (B) TYPE: nucleic acid
- 15 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

20 CTCAAGGATG TAAAGGCTCT GCAGATTTCG GGAGGCCTGT CTCCCAGCAC CTGATGGGAC	60
ACTTTTGCC CCACTGTAAA TTCTGGGTGT ATCCTCCACT GTATGCTGTC ACCCCAAGGG	120
25 CAAGCACTGC ATCTGCTTAG TGAAGGATTT ATTGTTCGGA AGATACATTT TCCCCTTAKAG	180
CAGAGAGTGG CGTATCCTGG CAGTCTTCGG TGAGCCAGTT GTACCAGGAT TATGAAATGC	240
30 AGATGTTTAC TGTCATTCATG TTGCTGTCAAT TGCTACTGAG GAGTACTGAC CAGAACATC	300
TGCAACTYTT AGTTGGCAGA GAGGACCACT ATGGCGGGTA GCTCTTTCT TTCCCTGCCAT	360
TGTGGGGATG ATTCCAGGCC AAAGATGATG GARAAGTATG GAAATCATCT GAAAGGTTGA	420
35 AGCTTGGCAC GTGAAGCCAT TCATGACTTT GTAAGGCAGT TTTGCTGAAG GCCAGTTCTG	480
CCCTGGGAGG GACGGAGGTG AATCCCTCTG AGTACCTGTG GTTTCTTAC TTCCCTGCTGA	540
40 ATTTACCTAA GTGCCTGTTG TTTGCTTGCT GTGGAGGCTT TCTGGTATTT CATTTCAGGT	600
GCAGATGCCT TCACTTTCCC ACCRAAAA CCCMACCAA ACCTAAGACC TTACTGCAAC	660
TAAGTYTNCC AAGTACTTTT TAACCAATG GGATGAACAG CCTGTGGTCT GCTCAGATCA	720
45 CCCTGAGTGC GTGTGAGAAG GCMINGGCTT TGCCAGGAAA TCCAGGAAGG CAGGGCGGGG	780
CTGTGTTGGA AGCTGGCTTA GCTGGTGGGG CAGCCTTATT TCAATTAAAA GGGCATTGAC	840
TGGGAGCAGC AGTCCTGGAG TTTGTTGCAT TTCTTATTGC CCTCAAAATG AGAAACCAGG	900
50 AAAATAGCAG ATGGAGCCT TCGAGAAGGC AGTAAATGGC TGTGTTTATT GACAAAAGGA	960
AAACATTAA CTGCCATCTC ACTGATGGCA TCTCACTGAC TTAAAATGAA GGCANGTTGT	1020
55 AGTAAAAAAA AAAGTCTACA TTTTCCACC GCCACGTTCT TATATCCCTGT TTGTCAGCCA	1080
CTGCTCANAA GGGCATGTTG TCTTGCGGAN TANAGGCCT CTCCCTCCCT CGTTTCCCT	1140
ATAGGTTGGG TG	1152

(2) INFORMATION FOR SEQ ID NO: 68:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2483 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

AGCAGGCGGT GCGCTGGGG CGGGAGCAGC GCGKAGCCCG GCTCGGCCAC ACCGATGCC	60
CGCCGCCATG GGCTCCTCGC AAAGCGTCGA GATCCCGGGC GGGGGCACCG AGGGCTACCA	120
CGTTCTGCGG GTACAAGAAA ATTCCCCAGG ACACAGAGCT GGTTTGGAGC CTTTCTTTGA	180
TTTTATTGTT TCTATTAATG GTTCAAGATT AAATAAAGAC AATGACACTC TTAAGGATCT	240
GCTGAAASCA AACGTTGAAA AGCCTGTAAA GATGCCTATC TATAGCAGCA AAACATTGGA	300
ACTGCGAGAG ACCTCAGTCA CACCAAGTAA CCTGTGGGC GCCCAGGGCT TATTGGGAGT	360
GAGCATTCGT TTCTGCAGCT TTGATGGGC AAAATGAAAAT GTTGGCATG TGCTGGAGGT	420
GGAATCAAAT TCTCCTGCAG CACTGGCAGG TCTTAGACCA CACAGTGATT ATATAATTGG	480
AGCAGATACA GTCATGAATG AGTCTGAAGA TCTATTCAAGC CTTATCGAAA CACATGAAGC	540
AAAACCATTG AAACTGTATG TGTACAACAC AGACACTGAT AACTGTCGAG AAGTGATTAT	600
TACACCAAAT TCTGCATGGG GTGGAGAAGG CAGCCTAGGA TGTGGCATTG GATATGGTTA	660
TTTGCATCGA ATACCTACAC GCCCATTGTA GGAAGGAAAG AAAATTCTTC TTCCAGGACA	720
AATGGCTGGT ACACCTATTA CACGTCTTAA AGATGGGTTT ACAGAGGTCC AGCTGTCCCTC	780
AGTTAACCCC CGCTCTTGT CACCACAGG AACTACAGGA ATTGAACAGA GTCTGACTGG	840
ACTTTCTATT AGCTCAACTC CACCAGCTGT CAGTAGTGGT CTCAGTACAG GTGTACCAAC	900
AGTACCGTTA TTGCCACCAC AAGTAAACCA GTCCCTCACT TCTGTGCCAC CAATGAATCC	960
AGCTACTACA TTACCAGGTC TGATGCCATT ACCAGCAGGA CTGCCAACCT TCCCCAACCT	1020
CAACCTAAC CTCCCAGCAC CACACATCAT GCCAGGGTT GGCTTACCAAG AACTTGTAAA	1080
CCCAGGTCTG CCACCTCTTC CTTCCATGCC TCCCCGAAAC TTACCTGGCA TTGCACCTCT	1140
CCCCCTGCCA TCCGAGTTCC TCCCGTCATT CCCCTGGTT CCAGAGAGCT CTTCTGCAGC	1200
AAGCTCAGGA GAGCTGCCTG CTTCCCTCCC GCCCACCAAG AACGCACCCCT CTGACCCCTGC	1260
CACAACACT GCAAAGGCAG ACGCTGCCTC CTCACACT GTGGATGTGA CGCCCCCAC	1320
TGCCAAGGCC CCCACCAACCG TTGAGGACAG AGTCGGCGAC TCCACCCAG TCAGCGAGAA	1380
GCCTGTTTCT GCGGCTGTGG ATGCCAATGC TTCTGAGTCA CCTTAACCTTT GAACCATTCT	1440

	TTGGAATTGG CGTGGTATAT TTAACCACGG GAGCGTGTCT GGAAACGCAA ACTATCAITA	1500
	ATTTCTACT AGTTTGTACC GTATCTGTAG GCATCCTGTA AATAATTCCA AGGGGAAAAC	1560
5	TAAACGAGGA CGTGGGTGT ATCCCTGCCAG GTTGAGTGGG GCTCACACGC TAGGGTGAGA	1620
	TGTCAAGAAAG CGCTTGTATT TTAAACAAACC AAAAAGAATT GTAAGGGTGG CTTGCTGCCA	1680
10	GGCTTGCACT GCCGPTCCTG GGGGTGTGCA TCTTCGGAA AGGTGGTGGC GGGGCGTCCA	1740
	CTAGGTTTCC TGTCCTCTGC TGCTCCCTTC GAAAGAAAAT GAAATATTCT ATGCCATAATA	1800
	CTCACACGCA ACATTTCTTG TACTTTGTAA GTCGTTGCG AGAATGCAGA CCACCTCACT	1860
15	AAACTGTAAA CGGTAAAGAG ATTTTTACTT TTGGTCTCCG TGAGTCGCAT CTCTACTAAG	1920
	GTTTACACAG GAATTCCACC TGAAGACTTG TGTTAAAGTT CTACAGCGCG CACTGTTAAC	1980
20	TGAACGTCTT TTTCTTCAGC CTATACGCGG ATCCTTGTCT TGAGCTCTCA GAATCACTCA	2040
	GACAACATTT TGTAACTGCT GCTGTTGCTT TCTACATACA CCTTATAAAAG TGACATTCA	2100
	AAAGAAATAA GGTGCCACAG TTTTAAACCA GAAGGTGGCA CTCTGTGGCT CCTTGTAGTA	2160
25	TTATAGCTAT ACTGGGAAAG CATAGATACA GCAATAAAAGT ACAGTAATTT TACTTTTTTT	2220
	CTTGTGTTAC ATCTAAATTAA CAACCCCTAA TTGCCACGTG TGCACTTACT ACTCTCCAGT	2280
30	ATGTCCTATT ACTCTCCAGT ATGTCACGCA TCTTTAACTT TTCACGTCT ATGTTTGCTT	2340
	TCTCCCATTT TTAAGAGATG GTAAGTTAAC TGGAATTGAT TTACTGAATG AAATTAAATG	2400
	CAGATATCCC TGTTTTGAA ATAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA	2460
35	AAAAAAAAA AAAAAAAA AAA	2483

40 (2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

50	GAGAAATGGA GCTTTGTAG ATAAAAATTT TTTCACCGCA AACAGTCATT TTCCAGTGAA	60
	AGGAGAGCGT ATCCGCCGTA GGATGGACTT AGATGTTGTA AAAGCTGAGG CCACCGAGGA	120
	TATAACCTCC GGGGTCTTTT GCCTCCCTTT CCTTAGACTC CCTCCAAACT CGTGTATCTT	180
55	TCCCTTCAGCA GTACTGGGCT CCACCGGAAC CTAGTCCTTT GTCTTTACCC TATTACCTTT	240
	CATAACATCC TAGTTGAAAA GTATTTATTTC AACCGCGTTT GAAAATGAGA ACAGGTTCAC	300
60	AGARGCTAGG TTACTTGCGA AGGTCGTTCA ATTAGTAACC AGTAACGCCA GGACTGCCAG	360

TTTCTTGCTT CCGAATTCTC ATGGTAGCTT TCACCARGCT CCCCGTCMAA TGCTAACGTC	420
AAACTACTGAA CTAGATTAGC AAAAAGGTCT TTTAACAGAA TTCCCTGGTTT TCAGAGAGAG	480
5 TTTCTTTCAT GAAGCGCCCC ATTCTACAG AGGAAAATAA ACTCCAAGCA GCCAGT	536

10

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

20 CCACCGGTCC GGCCCTTCTT GGCCAGAGGC GCGGGTTGGA CTCACGGGCG GGGCATGATG	60
GGTAACAGGA CCGGTGGGT CCCCAGGAAG TCCTAGAGGG GTGCGGGTT TGCGGTGGACA	120
25 AGCTTTCTTC GTCCTCTCCC GACAGAGCTG ACGTGTCTG GGTTCCACCG GGAGCGGGCA	180
TTTCCACCGG ACGGGAGGGT TCGGGGTGTC CGGGGCTGGG GAATACGTAG GGGTTGCCGC	240
30 GCGGTGTGGG GAGTTGGGC GTGTGGCTGC AGTCCCGGA GTTCTTGGAG GGGGTGGCC	300
CACCGAGCTT CGGGACCGGC TGATCTGCCG GTAGCTTGCC GGANGGARGG CGGAGCTGAC	360
TCTCCGTCCC TTCTCCCACG GCCTCCAGTG GTGGGTACGG GCACCTCGCT GGCGCTCTCC	420
35 TCCCTCTGT CCCTGCTGCT CTGGCTGGG ATGCAGATGT ACAGCCGTCA GCTGGCTCTCC	480
ACCGAGTGGC TCACCATCCA GGGCGGCCTG CTGGTTCGG GTCTCTTCGT GTTCTCGCTC	540
40 ACTGCCTTCA ATAATCTGGA GAATCTTGTCTTGGCAAAG GATTCGAAGC AAAGATCTTC	600
CCTGAGATTC TCCGTGCGCT CCTGTGGCT CTCTTGCAT CTGGCCTCAT CCACCGAGTC	660
TGTGTACCA CCTGCTTCAT CTTCCTCCATG GTGGTCTGT ACTACATCAA CAAGATCTCC	720
45 TCCACCCCTGT ACCAGGCAGC AGCTCCAGTC CTCACACCAG CCAAGGTAC AGGCAAGAGC	780
AAGAAGAGAA ACTGACCCCTG AATGTTCAAT AAAGTTGATT CTTTGTAAAA AAAAAAAAAAA	840
50 AAAAAAAAAA AAAAAAAAAA AAAAAA	865

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

	TCATCATATA CAAAGTTTT CGTCACACTG CAGGGTTGAA ACCAGAAGTT AGTTGCTTIG	60
5	AGAACATAAG GTCTTGTCGA AGAGGAGCCC TCGCTCTCTC GTTCCCTCTC GGCACCACCT	120
	GGATCTTGG GGTTCTCCAT GTTGTGCACG CATCAGTGGT TACAGCTTAC CTCTTCACAG	180
10	TCAGCAATGC TTTCAGGGG ATGTTCATTT TTTTATTCT GTGTGTTTA TCTAGAAAGA	240
	TTCAAGAAGA ATATTACAGA TTGTTCAAAA ATGTCCCCTG TTGTTTGGA TGTTTAAGGT	300
15	AAACATAGAG AATGGTGGAT AATTACAAC GCACAAAAAT AAAAATCCA AGCTGTGGAT	360
	GACCAATGTA TAAAAATGAC TCATCAAATT ATCCAATTAT TAACTACTAG ACAAAAAGTA	420
	TTTTAAATCA GTTTTCTGT TTATGCTATA GGAACGTGAG ATAATAAGGT AAAATTATGT	480
20	ATCATATAGA TATACTATGT TTTTCTATGT GAAATAGTTC TGTAAAAAT AGTATTGCAG	540
	ATATTGGAA AGTAATTGGT TTCTCAGGAG TGATATCACT GCACCCCAAGG AAAGATTTTC	600
25	TTTCTAACAC GAGAACTATA TGAATGTCCT GAAGGAAACC ACTGGCTTGA TATTCTGTG	660
	ACTCGTGTG CCTTTGAAAC TAGTCCCCTA CCACCTCGGT AATGAGCTCC ATTACAGAAA	720
	GTGGAACATA AGAGAATGAA GGGGCAGAAT ATCAAACAGT GAAAAGGGAA TGATAAGATG	780
30	TATTTTGAAT GAACTGTTTT TTCTGTAGAC TAGCTGAGAA ATTGTGACA TAAAATAAAG	840
	AATGAAAGAA ACACATTTA CCATTTAAAA AAAAAAAAAA ACTNGAGGGG GGCCCOGGTAC	900
	CCAAATGCC GCATAGTGAT CGTAAACAAT CT	932
35		

(2) INFORMATION FOR SEQ ID NO: 72:

40	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 996 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
45	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

	CGCCTGGCAC CATGAGGACG CCTGGGCCTC TGCCTGTGCT GCTGCTGCTC CTGGCGGGAG	60
50	CCCCCGCCGC CGGGCCCACT CCCCCGACCT GCTACTCCCCG CATGCGGGCC CTGAGCCAGG	120
	AGATCACCCG CGACTTCAAC CTCCCTGCAGG TCTCGGAGCC CTGGAGCCA TGTTGAGAT	180
55	ACCTGCCCCAG GCTGTACCTG GACATACACA ATTACTGTGT GCTGGACAAG CTGGGGACT	240
	TTGTTGGCCTC GCCCCCCGTGT TGGAAAGTGG CCCAGGTAGA TTCCTTGAAG GACAAAGCAC	300
60	GGAAGCTGTA CACCATCATG AACTCGTTCT GCAGGGAGAGA TTTGGTATTTC CTGTTGGATG	360

	ACTGCAATGC CTTGGAATAC CCAATCCCAG TGACTACGGT CCTGCCAGAT CGTCAGCGCT	420
	AAGGGAACTG AGACCAGAGA AAGAACCCAA GAGAACTAAA GTTATGTCAG CTACCCAGAC	480
5	TTAATGGGCC AGAGCCATGA CCCTCACAGG TCTTGTGTTA GTTGTATCTG AACTGTTAT	540
	GTATCTCTCT ACCTTCTGGA AAACAGGGCT GGTATTCTTA CCCNGGAACC TCCTTGAGC	600
10	ATAGAGTTAG CAACCATGCT TCTCATTCCC TTGACTCATG TCTTGGCAGG ATGGTTAGAT	660
	ACACAGCATG TTGATTGGT CACCTAAAAA GAAGAAAAGG ACTAACAAAGC TTCACTTTTA	720
	TGAACAACTA TTTTGAGAAC ATGCACAATA GTATGTTTT ATTACTGGTT TAATGGAGTA	780
15	ATGGTACTTT TATTCTCTCT TGATAGAAC CTGCTTACAT TTAACCAAGC TTCTATTATG	840
	CCTTTTCTA ACACAGACTT TCCTCACTGT CTTTCATTAA AAAAGAAATT AATGCTCTTA	900
20	AGATATATAT TTTAYGTAGT GCTGACAGGA CCCACTCTTT CATTGAAAGG TGATGAAAAT	960
	CAAATAAAGA ATCTCTTCAC ATGARAAAAA AAAAAA	996

25

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 785 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

35	GGCACGAGGG GCTTTGCGTA CACAATAGCT GCTAGGAGTA CCCAAAGCCT GARTACARCC	60
	TGCTGGTGTGTC ATGGCCACGT GTGAGCAGGC CAGCGTCAMA CGGCTCGCTG TGACCCGTCC	120
40	CGRAGACTGA AATGGGCCTG GGTCTTCTCC TKGTCTGTG ATWAAAGTCC TCTCTTGAAA	180
	GTGGAGAGCA AAGGCACACA GAGGTGCGCG CTCACAAGAA TTCCCTCCGG TGACTGGGTA	240
45	ATCAATGTTA CTGCTGTTTC CTTTGCAGGA AAGACCACAG CAAGATTCTT TCATTCTGTCT	300
	CCTCCTAGCC TGGGGGACCA GCCTCGAACT GACCCCTGGAC ATCAAAGGAG GGATTATGTG	360
	GCTGCTAAAG CCATCGGCC ACAGCCCTGT TCACRTCTTG GTGCTTCTCT TTCCCAGAGG	420
50	CTGGTCCCAG CCAGGCACAC ACAAAAGGCA GATTCTCGTA AACSCAGCCT CCCTCCCTGG	480
	AGGCTGCCTC CTGCCCTGGA TCTGGAGTGG AGCTGCTCTG AGATTTTGAG TTCTTCTGCA	540
	GAGATGATTA AATATATCCA AGAGACATTG GAAAACCTGC TGAACATTAA ACATTGGTCT	600
55	GCTCAGCACA TGGCTGGATG CGGATATTTC TATAATTCCA GAAAGTCACA CAGCTCCTCT	660
	GTATGAGACC AGTGGGCGCC ATTAAAAGA ACAGGATGAG AATCTAAGAT ATATTATCAA	720
60	TAAATGTAAT GGATTTTTT TTGTAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	780

AAAAA

785

5

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1069 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TCCTCACCAT	TCCCCTAGGN	CAGGTCCCTG	CAGGTCCCCAC	ACTTCTCCCA	GGTCCCTAAA	60
CTTGGGTCGG	TCCTTTCCCT	GGAGTAGCTG	GNTCCTCCAG	TCGAGGTCCC	TGTTCAGTCG	120
20 GTTCTTAGGC	TCCTGCACAT	GAAGGTGTGT	GCCTGTGGTG	TGTGGGCTGC	TCTAGGAGCA	180
GATACTGGCT	GGTATAGAGG	ATGCAGAAAG	GTAGGGCAGT	ATGTTTAAGT	CCAGACTTGG	240
25 CACATGGCTA	GGGATACTGC	TCACTAGCTG	TGGAGGTCTT	CAGGAGTGGA	GACAATGAGT	300
AGGAGGGCAG	AAGCTTCCAT	TTTTGTCCTT	CCTAAGACCC	TGTTATTGTT	GTTATTTCT	360
30 GCCTTTCCGA	GTCCTGCAGT	GGGCTGCCCT	GTACCCGTAA	CCTCATGAGC	CTCTAAGGGA	420
AAGGAGGAAC	AATTAGGACG	TGGCAATGAG	ACCTGGCAGG	GCAGARTACA	AGCCCAGCAC	480
CAGTGTCCCA	GCCTTACTGG	GTCCTTACCC	TGGGCCAAC	AGGGAGGGCT	GATACTCCCT	540
35 TGCTCTTCCT	AGATGCCAC	CTCCTACAAT	CTCAGCCCAC	AAGTCCCTTC	CACCTAGGG	600
GGCTTGCTGC	ATGGCAATAA	CTCATAATCT	GATTTGGAGG	TTTGCCTTT	ACAGGGCAG	660
40 ATTTTCTGCT	CAGTTCAACA	ATGAAATGAA	GAGGAACCTCC	CTCTTTCTAC	AGCTCACTTC	720
TATCAGAGGC	CCAGGTGCCT	CAGAGCCACA	TTGAGTTGCT	TTTCTGGGA	TGAGGAAGTA	780
GGGTTAAACT	CCCCAGTTTC	CTGAGGGAGG	CTCCTGACAG	GTCCCCTTTG	TCAGACCCCTA	840
45 CCACAGCCTG	GATAGGCAGC	CACATTGGTC	CTCGCCCTTG	CTCGGNACTC	CGTGGTGGTC	900
CTGCCCTTCT	CCCTGCATGC	CTGTGGGTCT	GCTCTGGTGT	GTGAAGGTCG	GTGGGTTAAC	960
TGTGTGCCTA	CTGAACCTGG	CAAATAAAC	TCACCCCTGCA	AAGCCAAAAA	AAAAAAAAAA	1020
50 AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	1069

55

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 831 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

5	GGACATTAGA TCACTGTGGA CCTAAAACAA ACAAAACAAT ATAAGGAAAA TGGCATTAGA	60
	AATGGTCTGG GGATCAGTTT ATCACTGCAG TIGTTACATC ACCCCATGGT CTAAAATACA	120
10	GAGCTTTAGT CTGTCTCTGT TTTCAGTTCAT TTTACAGGAG GTGAACATCA CACTTCCAGA	180
	AAACTCTGTC TGGTATGAAA GGTATAAATT TGATATTCCCT GTCTTCACT TGAATGGCCA	240
15	GTTTCTGATG ATGCATCGAG TAAACACCTC AAAACTTGAA AACAGCTCC TGAAACTTGAA	300
	GCACCAAAGT ACTGGARGCT GACTGATGCC CTCATGATTT TCCACCCCTCT CTTCCCATAA	360
	AGCATCTTCC TAAGGAAATG AMCATGGCCT GATACTCATT TTGTCACTTG TACAGAGCCC	420
20	TAAGGATGTT CTGAATTCAAG TGGTGCCAAA TAAATGTTGA CATTCCCCTT TTGGTTGATG	480
	GAAGTATCAG TGTGGGAACG GTTTCCTTAA TGGCATTITTA TAAAATAAKA AKAKCATATT	540
25	AGCAGGGAGG GAGATGATGG AGGGAGGGAG AAGTCCATTG GTCTTATTTA TCCTTTTGT	600
	ATTAATAGAG AAGCACTICA CAGTCACTGG CAATGCCATT TATAGGAAGA AGGTTCTGCA	660
	TTCCTGCTGC TCCCGGAGGG CTTAACTTTT TAATGAAAGA ATAAATGCTC TTCCACTCAG	720
30	TAGATAAAAGT GAAATGTGAA TTGTTAATAA CTGTGCACGG TCAATAAAGC GATGTTTAA	780
	GGAATACAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAACTCG A	831

35

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

40	(A) LENGTH: 590 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
----	----------------------------------------------------------------------------------------------------------

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

45	TATATATAGA CNGTTAATAG TCGTGANGN TGTGNACGAA CATTAACCGA AGTACCATGT	60
	AGCCAGTCGA ATAACNTATA AGGACAAAGT GGAGTCCACG CGTGGGGCCG TCTAGACTAG	120
50	TGGATCCCCC GGCTGCACGA TTCCGCACGA GCTGCCAGGT GAGGAGCAGA GAGACTGTC	180
	CCTTGGGTGG AGAGGTGTGG GCATGAGAGC CACCCATTGC CAAGCAGCAA GAATGTTCGT	240
55	GCTTTTTTCC CTTCCAAAAT ATGCAGGGCT CAGGCTCCCA ATTCCGGGCC TGTCTGCTTT	300
	GCTTGTGTTT CTCCCTGCTCC TGTCTCCCG GAGGGCCAG GTGGAACATCA CGACAGGGAG	360
60	GGAGACGCTT CCCAAAAACC TGCAGGGCTA TTTCCAGAA TTTGGTTTC AAGTACAAA	420

CTTTTGTCC TGTAAAGATAT ATGCAGCCTC ACAGAAGCAG CCTCTGCCTC CACTTTACCA	480
GCTACGTTTT TATCTTAAGC ACATGGGGCT CCCTTAGAAC TTACTCCACT GATTAAAAAA	540
5 AAAAAAAAAA AAAACTCGAGG GGGGGCCCGG TACCCATTG CCCTAAAAGT	590

10 (2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1274 base pairs
- (B) TYPE: nucleic acid
- 15 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

20 GAGCCACCAC ACCTGGCCTG GAAGGAACCT CTTAAAATCA GTTTACGTCT TGTATTTTGT	60
TCTGTGATGG AGGACACTGG AGAGAGTTGC TATTCCAGTC AATCATGTGAG AGTCACIGGA	120
25 CTCTGAAAAT CCTATTGGTT CCTTTATTTT ATTGTGAGTTT AGAGTTCCCT TCTGGGTTTG	180
TATTAATGCTT GGCAAATGAC CTGGGTTATC ACTTTTCCTC CAGGGTTAGA TCATAGATCT	240
TGGAAACTCC TTAGAGAGCA TTTGCTCCT ACCAAGGATC AGATACTGGA GCCCCACATA	300
30 ATAGATTTCATTTCA TTTCACTCTA GCCTACATAG AGCTTTCTGT TGCCTGCTCT TGCCATGCAC	360
TIGTGCCTTG ATTACACACT TGACAGTACC AGGAGACAAA TGACTTACAG ATCCCCCGAC	420
35 ATGCCTCTTC CCCTTGGCAA GCTCAGTTGC CCTGATAGTA GCAATTTCT GTTTCTGATG	480
TACCTTTTTT CTCTTCTTCT TTGATCAGC CAATTCCCAG AATTCCCCA GGCAATTGTT	540
AGAGGACCTT TTGGGGTCC TATATGAGCC ATGCTCTCAA AGCTTTAAA CCTCCCTGCT	600
40 CTCCCTACAAT ATTCAAGTACA TGACCACTGT CATCCTAGAA GGCTTCTGAA AAGAGGGCA	660
AGAGCCACTC TGCGCCACAA AGGTTGGGTT CCATCTCTC TCCGAGGTTG TGAAAGTTT	720
45 CAAATTGTAC TAATAGGSTG GGGCCCTGAC TTGGCTGTGG GCTTTGGGAG GGGTAAGCTG	780
CTTTCTAGAT CTCTCCAGT GAGGCATGGA GGTGTTCTG AATTGTCT ACCTCACAGG	840
GATGTTGTGA GGCTTGAAAAA GGTCAAAAAA TGATGGCCC TTGAGCTCTT TGTAAGAAAG	900
50 GTAGATGAAA TATCGGATGT AATCTGAAAA AAAGATAAAA TGTGACTTCC CCTGCTCTGT	960
GCAGCAGTCG GGCTGGATGC TCTGTGGCCT TTCTTGGTC CTCATGCCAC CCCACAGCTC	1020
55 CCAGGAACCT TGAAGCCAAT CTGGGGACT TTCAGATGTT TGACAAAGAG GTACCAGGCA	1080
AACTTCTGTC TACACATGCC CTGAATGAAT TGCTAAATTT CAAAGGAAAT GGACCCCTGCT	1140
TTTAAGGATG TACAAAAGTA TGTCTGCATC GATGTCTGTA CTGTAATTT CTAATTTATC	1200
60 ACTGTACAAA GAAAACCCCT TGCTATTTAA TTTGTATTA AAGGAAAATA AAGTTTTGTT	1260

TGTAAAAAAA AAAA

1274

5

(2) INFORMATION FOR SEQ ID NO: 78:

- 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1133 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AGGATTTTC CTTGTTCAAC CAAAATCTGA GCATTCTTC TATGTTGAAA ACACTGAAA	60
ACTAATTIWA GTTAATGAAC TAGAAAGAAT ATTGATTIIW AAGAACAGA AAAATACTAC	120
20 TTATTTCCCT TCTCAAATAA CGTTTCTTTC AAAAACCTCT GGCTGAAGTA TAACATGCTG	180
GTAGTTAACCA TAAATCTTGT CTTCTCTTG TTCCTTATCT TTCTTGTGTT TTTAGATGCT	240
25 TGTATAAATG TCTTTGTTT TTATTAAGTG CCTAATTGAC AGAGCTTAAT TTGAAGAAGT	300
GCCCTAATT ATTGACCACT TAAGAATTGC CTTTATTGGG GTATTTATT TGTTCCCGCG	360
30 TCTTTTGAT GTTGTTCAGT CTACTCATCC CTGTGAGTAT GTGTGGGGGA CAGCTGATAG	420
AAGGGAGGAG AGTGTGTCTA TGTCAGGAT TGCCCTTAG CCACTCAGCC AGAGATCCAC	480
AGGGAGCAAC AAGGACAGTT TCACATGCTT AGACTTCTT GGAAGAAACA GTGAGGAGGA	540
35 GTAAGTCGTG AGTAGTGTCA AGCTGGATGT AGAATTGTCC TAAGGCAGTT GACCCCACCT	600
TCCAACATGT TTTCACTTTA TTTGCCCTC CCTACATTG GGTAGGTTTC CATTGGATT	660
40 TGCAGCAATA ATGACTTTAT TTCTCTCTTG GTCAGGATT GGCACATAAA ATCCTTTAT	720
TATAGAACTA GCTATTTAG TTACATAGTA ATGTAACCAA TGGAGAGATT TATAGAGAAT	780
TTTGKTTTG CTGTCATATA TGTCATTTT GGAGACAGAT ATGATAGAAC TAGAAATTAA	840
45 GTTGCATTTC TGCAAGTGCC ATTTGAATGA ACTTCAAGTA TCTTCTTAAT TATTAATT	900
TCTGATGAAG GCATTGTAAC AAATATATAG TATTATCAA TCTAATTAAT ATTTGAAAT	960
50 ATTAATAAT AGGTATTTA TTTACTGTAA AAAGTCAAAC TTCATTATGT AGATAAAATCT	1020
TATTCTTTTC ATTCTTCCCC CTGTTTACAT CCTTTTACA AAGCTTAGTC ACCAATTAAA	1080
GCTTCCAT CAAAAAAA AAAAAAAA ACTCGAGACT AGTTCTCTCT CCT	1133

55

(2) INFORMATION FOR SEQ ID NO: 79:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 661 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GAATTCGGCA CGAGGGAAA AGGATGCTGA ACGAGAGCAG AAAGCCCTTT TCCTTTCGCTT	60
10 CACGCCCTTC CAGTCCTTAT TTTAAACTCG GGTTCCCTTT CTGTGGTCGC AGCAACCTTT	120
ACTCCACCTG CACTGCTGCT CCTGGGGCT CCCCAGGCCT CCCTCTGCCT TTCTACCCAG	180
15 TGGCTGACGG GATGCCTGTC TTGCCTGGAC GCACCACTGC TCTCCGTGCC CTCACCTTGG	240
CTTTTGCTGT GGCCTGCTCT GGGGTTGAAG CTGGCCCCATG TGTCCTCCGG AGTCATGGCT	300
GCTCCCTCTG GGAGGCCTCT GTGTGCGTCA CGTCTTCCAC ACCTGGGGC AGCTGGCGAG	360
20 CCCGTGCTCT GTTCCCCCTCG GCTGCTTGGC ACAGAGYTGC AGCCTGGAY TCTCCGTGGA	420
CCCAGACTGG GGATTTTGCC AGGGGGGCGA TGGGAGGAGC AGGTGCTTTG CCTGGCGGCT	480
25 GTGTCTGCAT TTCTGGACGC CCCAGAGCAC AGAACATTGCC GGCACCTTGA GGTCTTCCTC	540
GGCATGTGCC AGATTACATG AGTGACGGCT GGGAAATATGT TTTCTTTTTT GTAATGGAGG	600
CGTGTTCAC ATATAGTAAA GCTCACCAAA AAGTAAAAAA AAAAAAAA AAAAATCTCG	660
30 A	661

35 (2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1378 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

45 ATGGGTACC GGGCCCCCCC TCGAAGTTTT TTTTTTTTTT TTTTAATGAA AGCTCTCAA	60
TAAGCGATT TATTCCATAC CATGATTGCA GACATTTACA AAACCATAAC ATCTGAGTTC	120
50 ACCTAAAAAA ATAACCTATA TAAAGCAGTG ATATACACAG CACAAAATAG TTCAGGGAGG	180
GGGCAGGAGC AACTTGTAAAT AATAAAAATG TAAACGTGAA AAAAAGGATG GAATAAAAGT	240
CCCTACTTAT TTCTACTTAA GATGTCAATGT GATAATATTT TACAATGTCC TGTGGGTCAA	300
55 TGTATGTATG TGTATATGTC TGTATAACAT ACACATATAAC AGTACATTCT CTTTCCCACA	360
CATATACATA CACACATAAT TATTTGCAGT TCAGTTTAGG GCAATTCTAA TATGCCACTC	420
60 CGTACAGTTG TTTGAATCAC ATTGGACCC GCTTTCTCA CAAAAGAGGG GAGAGAGCAG	480

	GAAATAAAAA GGTTGGTTTG GTGTGACTGA GATTCCCTTG TTTAACTGTA CACTGTGATG	540
	AATAATTTTC TTCCGTAGTA GTTCTGTGAA GGGCTGACTC ACTGTGGTT TCATGAGGAG	600
5	ACTTGGTAAT GGATCACACG CTCATTGTCA TGCTAGGGGA GTAACCTCTCA CTCTGAAAAG	660
	GATTAAAGAA ATTTCCCCCC ATTTCGCCAT CATCCCTTGG AGTGCCCGGT TGATTACTCA	720
10	GGCTCATATT ATTGGGAGAA TTCTTGAAA TACTGTCCAT ATCTCCTGAG CCTAAAGAGC	780
	CATTCATGTG ATGTGACTCC ATTCCTCTA ATCCACCCAT GGGACCATCT GACCCAGGR	840
	CCATTGGAAA ATTAGGTCTG TTAGGTCCAG GAGGTACTGC ATTCAATTAA GTATACATGT	900
15	TATCACCAGA GTTGGTIGAA TCTGCTGGAC TAGGCATGAT GGGTGTTCCT GGTGGCCCTC	960
	CACCTCCTGG AGGACCTACA TAATTCCCAG GAGATGCTGA GGAGTATGGT ATTGAATTGG	1020
20	CATTTGGTGG GTTTGGCAA GGCTTACAC CACCTGGACC CATGTCATT CCAGGCATTC	1080
	CAGGGCCACC TAAAGCATTG AGTGGGGTC TCATTGCACC TCCATAGTTC TGTGGTCCTA	1140
	AGGGCACCAT TCCCTCTGGGA GGAGTCATTTC TCTGCATTGG CCCACCCATA TTTGGATGTC	1200
25	CTTGTGTGCG AGTTGGATCC ATTCCACTGG GGAGTAATGG CTGACTTCCT GGGACACCTC	1260
	CAAGTGCCTG ATTAGGTATC CTCAATGGGG GCCTTGGACC TCCAGGGTAC CGAGGTGACA	1320
	TAAAAGGGTA ATCATGGAAG GCTTTGCTT CACTTGAGTG TTCACATGTT TCACGTCT	1378
30		

(2) INFORMATION FOR SEQ ID NO: 81:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

45	ACTTGTCCA AATGTGTCTG TCACATGTAG TCAGCTGNAG NAATTAAAAA TGAATTGCCA	60
	AGTGAAGAGT CTGTGGATTA ATTGGCCGTT AATTAACAGG CTTTATCAAT GTGTCTCAA	120
	GGGAGAGGCC CAACCTTAAT TAAGGAGCTA AACTTCCTGA GTGAGGGCT GTGAGGATGG	180
50	AGGTGGAGGA GCCATCTGGG GCGCGTGGTC GCGGGGCCAG CAGATGGCGC CTCCCTGGCT	240
	GAGCTGCCCG CACCGCCAGT TCCCTCATT CCACTCAGGA AGGCAGAGAA GGCAGAGTGA	300
	TCTCCTCAAG GAAGAGCTTC CCCAGCCPTC GGGAGCAGCT GCCAGGGCT CGGGAAATAA	360
55	GCCCTACACG CGCGCGCCTG CCTCCAACTC ACTAACCCCTG CGCTCTTGT CTTTCAGATT	420
	CAACGCGTTC AACAGAAGCC ATCCCCAGCC CAGCTTAAAT TATAAAGATA GACAATAACT	480
60	CTGTTCCAAT CTGCGTGGTG CTTCTTTAGT AAATACTGTA CAGATTTAC CATGGAGAAC	540

	TTTTTTTTTA GTTTTTACCT TTTCTTAATT ACCCTTATTC CGAATGGACG AACACTTTCT	600
5	ACCACTGCTG ACCATTGTA AATAACCGTGT ATATAAATCC CATTGAAATA ATGCCCTGGA	660
	ATAGAACATC TCAAATGCTG CTTAATTACA GACTCAGGTC GATTACTTGT ATTCATGTA	720
	ATGTTCCCTCC AAGTTAGACA TCIGGGCAA GACCAACCGG GAGACCATGG AATTGTCAAA	780
10	AGTACAAACT GACAGTGTGT ATATTTAATT TAAAGACTTA TTTAAAAACT CACAAGCTCT	840
	CACCTAGACT TTGGAGAGCA GTCIGTTTC TGTAATGTCT GATACTAGAA ACTAATTTGC	900
15	TTATTTAGT TGTATTCAAG ATTTGAAGAT GTATTTATA GACAAGTTCT GTTTTGAAAC	960
	TTTGTGGAAC TGTTCCAATC AATCAATTTC CCAGTTATGA TGAGTATTAA CATTATGAAT	1020
	GTATAACCCA GACATGATTT GTAAAGCCGA CAGTATGTTT CTATTACACA ACACCTTTTG	1080
20	ATACAGCGTC TCTTGTCTTC ACTGATACTG GAGTCTCCGT TGTCTGCNNG GTCCCTTCGA	1140
	GTTCTAGTT ACAGACACAA TCATACTGTG ATTTTATTT TAATATGGAT ATGCTATCAA	1200
25	ACTGTGATAC ACITTATAATT CACTGGTCCT GCATCAGGAG ATGGAGTGGG GAAAACGTGA	1260
	TTTAATACAG TTTGTATCTG AATAATCTGT ATGGTTTATA CAGTTGTGT TGTCAGAGA	1320
	TGTTTAAAGT TTGATCTTGT TTTTCTAAA GATTAAAAAA GCACCTGCC CACTGTAAAT	1380
30	ATACAGCATG TAAAATTCT RTAGTATATA AATGGCAGCA AATCACAAAA AAAAAAAAAN	1440

35 (2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1381 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

45	CCCGGGCTGC AGGAATTGK YACGAGGCCA GCAGTTGCTC CCAGTTCAAGG AGGTGCTCCT	60
	GTACCCCTGGC CACAGCCAA TCCTGCCACT GCTGACATCT GGGGAGACTT TACCAAATCT	120
50	ACAGGATCAA CTTOCAGCCA GACCCAGCCA GGCACAGGCT GGGTCCAGTT CTGACCTGAG	180
	CACGGTTTTT CCTCATGTGA CTCTGGAA GGCGCTCCCT CATCTGGGCC AAAGGAAGGA	240
	GGACGAAAGCC CTCCCTCAGCT GGCCCTGTGTT TGGGGCATGA ATCTCTOCTC TCCCTCCTGT	300
55	CTGGCTCTGT TGACAAACCG GGCATGTTTG GCAGTAAATT GGCAACCGTGT CACACTGTTT	360
	CCTGGGATTC AAGTATGCAA CCAGAACACA GGAGAAGAAA AGCTCCAGGA TCCCTGTC	420
60	CATCTGTCCT CTTGATGTGA GAGAGACTCT GAGACTTCCTT CCATCGCAAT GACCTGTATT	480

	AAACACAAGC CCCCCAAGCA AAAGAAGAGG TTGAGTTTGC TGCCAGGATT CAGATCAGCC	540
	CTTCCCAGGG TCTGCAGGTG TCACATGATC ACAGITCAGC GGGAGGCTTT CCGTACCCAC	600
5	ACTGGCTGTA GCACITCAGT CCATCTGCC C TCCAGAGGAG GGTITCTTCC TGATTTTAG	660
	CAGGTTAGA GGCTGCAGCT TGAGCTACAA TCAGGAGGGA AATTGGAAGG ATTAGCAGCT	720
10	TTTAAAAATG TTTAAATATT TTGCTTTGCT AATGTGCTGA TCCGCCTAA CTCATCTTG	780
	CAAAAGGAAC TGCTCCCTCG GCGTGCCCCA GCTGGGGCCT CTGAAGGGAT TCCTCACTGT	840
	GGGCAGCTGC CCTGAGCTTC AGGCAGCAGT GTTCATCTCT GGCCAGTTGT CTGGTTCCA	900
15	TGTATTCTAG GCCAGGTAGG CAACACAGAG CCAAGGCGGG TGCTGGAAGC CAGACGGAAC	960
	AGTGTGGGG CAGGAAGGTG GATGCTGTTG TCATGGAGCT GTGGGAGTTG GCACTCTGTC	1020
20	TGCTGGTGGC CCTCTCGGCT CACATGTTCA CAGTGCAGCT CCTGGCAGAC TTGGGTTTC	1080
	TCTTTGGTGG TTTCTAAAGT GCCTTATCTG CAAACAACCTT CTTTCTCCT TCAGGAAC	1140
	TGAATGGCTA GAAGAAGGAG CTCAGTAAAC TAGAAGTCCA GGGTTGCTTG GTTTACTGGT	1200
25	TTATAAGAAA TCTGAAAGCA CCTCTGACAT TCCCTTTATT AACTCACCTC TCAGTTGAAA	1260
	GATTCTTCT TTGAAAGGTC AAGACCGTGA ACTGAAAAAA GTGTTGGCCT TTTTGCGGGAA	1320
30	CCAGATTTTT AAGATAAAAT AAAATTTTTT ACTTCTGTCA AAAAAAAAAA AAAAAAATNT	1380
	C	1381

35

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1706 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

	ACTGCACCAC TGCCCAGGTC TCCCGGCTGG ATGAAGACGT GGTCATGAG GAAGCTGGCT	60
	AGCTCAGACT GGAGAGTAGC TTCAGGAAAA AAGACAAGTG GCCTAAGGAA ATCACGGCCC	120
50	CCAACTATCA TCTGAGGGCT AAAGATGAGA AGTAGATCAC TTAATAAGAC AAAACCTGT	180
	AGGGGGAAAA GAAAGGATGT TTAAAAGGAC AGAATGTTTC CCAAGGTAGA AATGACACTG	240
55	TCAATTCTC CTTGGAATGG GGGCAGGGAT ACTCGCTTGC TTGCTCCCAC TTGAGTCAGT	300
	ACTCACCTGC TCCCTGGATCT CAGTATCCAC ATCTGAGAGG CAACTCTGGC AGAGITCACA	360
	GAAGGCCACC ATTCTGTCCC TCAAACCTCGA CAGCTGCTTC TGTGGGCACA GTGGCTTGAA	420
60	GGGGAAGAAT GAAGACACAG ACTCCTCTGT TCCCATTATC CCATCTAAGA CCCACACTCA	480

	CCTGGGAAG CATCTGATTT AGAAATGTGG GTTAGTGTCC AGAGAATGGA AAAATAGACA	540
5	AGAGTCAGG CTGGCAGGAT AACCTGTAAC AACAAAGGGT TTGAAAAATG AGGTTGGT	600
	TAGGAGAGGG AGAGACAGAT ACCAGAAC ACACCAGTGA AGAGGAGAGA AAATGAGTAA	660
	AGGGAGAGCT AATTCTTTT CCAGTGGAAA ATGAGTGATA TTCTGGACAT TCTTCAGAGG	720
10	CATCTACACG AAGTAGAAAT GTCACCGCTC CCTAATTAC TCTACGTCTT CTAGAATCCC	780
	TCAATATTAT CCTTGGCTTC CAGGAAATCC AAGAAGACCC TGGAAGTAGA GTCCACCTTC	840
15	TAAGAGAGGA ATGTAAGAGG TGACCCCCAC CCACCTGATC TTCCCTCGCTT TGTCCACTCC	900
	ACCCACTGAG ACTTGACACA CCTAGTGGCC ACCTAGAACG TAGGTCTTA AAATYTAGCC	960
	CCCCAGCCCC CAACCCATCT CTAGCCTGTC CACTCACCTG GTGAGGAACY TYTCCITGNT	1020
20	CCACAGCYTT CTGCAGGAGT TGGCAACATG GCTCATAGAG CTCCCAGCGA GTCAGGTCA	1080
	GAGTGCTTIG GGGGAGAAAG GGGAAATGTTA TACTGGAAA GAACAGAGGG AACCAACTCC	1140
	ACAGACACCA GTAAAAACGG GATGGGAAG AGGAGGAAAG CCACTCACCTT GTAGAAGGCA	1200
25	GAGAGGCCTT TCAGAGTGGC TGCCAGATTA TATACCTCAT CCTCATCTAG GAAGGACGAC	1260
	TGAGAAGGAA AGAAGATCCA CAATAGCATT TCCCCAGAA CTCATCAGTC CACATCCCC	1320
30	GTCCTGCAGC CCCTCCCACC CTTGTTGGG GTGTCCCATT GTCCAGCCCC AGCTCTACC	1380
	TGTAACAGCT CTTCAAGCTC CTGCTGGAAR CGGTCACTCA GCAAATCTAC TAGCTGGCTG	1440
35	CGGGCAAAGT CGGCCCCGCT GAAGAAAGTG AATTGGGAT TACAGAGCAG GTAAGAGCAT	1500
	GOGCCCCAGC CTCAAGCACC GCTGGCTCTG CATGCTTCAC CACCACCTCC TGGAGTTGCT	1560
	GCAGGAACAG CTCCAGGTGC TGAGAAGAAA AGGCAGAAGA TGGTGTGCTG TGGGGATGGG	1620
40	AGGAGGACAC TCTTCTGGCG CGAAGTGGAA CGGGGTTAAA ACCATTAAC TTCAAGGATA	1680
	AGATGCCTAA RAAAAAAA AAAAAA	1706

45

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

50	(A) LENGTH: 573 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

	GAATTGGCA CGAGCTTGGT AGCCTTAGAA CTGCATGAGC TGCTTTACCA CTGGAAACA	60
60	CGAGCACAGC CTAGCTTGAT TTGTATGTG GTATCAGATC TAAGGTGGAT GGAATTCAAGG	120

	ACTTCCIGTC TACTCTTGA TTTGTTTTA TTTTAGAAA TGTTTATTT TGTTTTATIC	180
	ATTTATTCACT CTTCAGAGAC ATGGTCTGGC TCTGTTGCC AGGATGGAGT GCATGGTGTG	240
5	ATCATAGGCC ACTGCAGTGT TGAGCTCCCG GGCTCAGGCG ATCCTCCTGC CTCAGCTYCC	300
	TTAGTAGCTG GGACTATAGG CACATGCCCT ACCATGCCCTG GCTTGTCTA CTTTTGAAAT	360
	GATGTCYCAA ACTAGAAGGT CTATTAATTT AAAAAATTAA GGATAGCATG CCATAATTAA	420
10	AAATAATAAC AGTGGGAAAA GGCACCTTCC AATGATTCAAG ACATCAACTT GTGATTTAAA	480
	AAAACGAAAA ATAATAATAA GGAAAAAAAG GGGAAAAGT TAAATAAAAA TAAAATTAAA	540
15	AAAAAAAAAA AAAAATCGA GGGGGGCCCG GTA	573

20 (2) INFORMATION FOR SEQ ID NO: 85:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 684 base pairs
 - (B) TYPE: nucleic acid
 - 25 (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30	CTCTTGGCT GTGTCTACCT CCTTCATCTG CTGCGCCGAC ATAAGCACCG CCCTGCCCT	60
	AGGCTCCAGC CGTCCCCCAC CAGCCCCAG GCACCGAGAG CACGAGCATG GGCACCAAGC	120
35	CAGGCCTCCC AGGCTGCTCT YCACGTCCT TATGCCACTA TCAACACCAG CTGCYGCCA	180
	GCTACTTTGG ACACAGCTCA CCCCATGGG GGGCCGTCTT GGTGGGGTC ACTCCCCACC	240
	CACGCTGCAC ACCGGCCCCA GGGCCCTGCC CCCTGGCCT CCACACCCAT CCCTGCACGT	300
40	GGCAGCTTIG TCTCTGTTGA GAATGGACTC TACGCTCAGG CAGGGAGAR GCCTCCTCAC	360
	ACTGGTCCCC GCCTCACTCT TTTCCTGAC CCTCGGGGC CCAGGGCCAT GGAAGGACCC	420
45	TTAGGAGTTG GATGAGAGAG ACCATGAGGC CACTGGCTT TCCCCCTCCC AGGCCTCCTG	480
	GGTGTCATCC CCTTACTTTA ATTCTGGGC CTCCAATAAG TGTCCTCATAG GTGTCTGCC	540
	AGGCCACCT GCTGCGGATG TGGTCTGTGT GCGTGTGTGG GCACAGGTGT GAGTGTGTGA	600
50	GTGACAGTTA CCCCATTTCA GTCATTTCCT GCTGCAACTA AGTCAGCAAC ACAGTTCTC	660
	TGAAAAAAA AAAAAAAA AAAC	684

55

(2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- 60 (A) LENGTH: 1036 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

	TGGAGGCAGA TGCACAGGAG AAAGGTTCCC GTCOGCACCC TCTCAGACCT GAGGCTGAGC	60
10	TTGCAGTGAG GGCTTCTCCT CGGCCCCCTCG CCCGCCCA GAGCTGCCAT CCCTGCTGTT	120
	ACAAGCCAGA GGAGCCCGGA TGAGGCCCC CAGATCACCT CCAGGGACTT GGGGTTCCCA	180
	TCTGAAATCC TTTATTTTG TACCATGGGG TGGGCCCCGG GCTGAGAAGG AAGAACACC	240
15	CTCTCCCCGG CCTCCCTCTGT CTGCACCCGT GGGGCTGTGA CTTACTCTG CCTCCAGGGG	300
	CGGGGGGGGGG CCCCTGGGA CCTCTTAAGG CCCAAGGTGG GCCCCAGGAC CTYTGCGAG	360
20	AGTGGAYTGC TCATGGCAGA TGTGTGGCAA TGTCTGGCTG WGTCTTCCG GCAMCTGGT	420
	YCCCTYTCCTT GGGYTCCTCT GCTGCATGGT GGATGTGCTC CTTCTGGCC CGGTACATT	480
	GCCTCCTTGA GCCTTAGTCC AGGGGGTCAC TYCTCCCACC CCACCTACCT CACAGGGTTG	540
25	TTGTGAGGGT GCACAGAGGA GCAAAGTCCT TGAAGGCCCT CAGGCAGTAT ATAGGGGCCG	600
	CCCACCTTCA GCTGCCCTGG GATGGGAAGG ACCCAGCCCG ACCCCTGGC ATAACACTGT	660
30	GTTTCAAAT GGAGATTCA GATATGGGA TGCAGGTGT GGGGAGCTGG CCTGGCAGAG	720
	TAGGGGTAGT TGGCTTGGCC TTCTCTTGG TGATCCCACC CCCAGCCATT TGCATTGCTG	780
	GCCCAGCGCC TGGCTGGGG CGCGGGAGA GGCAGCAGAA GGGCTGGC AGGGGCGGTG	840
35	GAGGACTCAG GAACTGCCCG GGGAGAGTGG GTATGGCGGC TGAGCCAGGG GCCCTCCTGT	900
	GTTTGACTTC CCGGGATGGG TCCPTGCTTC TCAGCTGTGT CCGACCCAC CATGTAATAA	960
40	AACCCAAAGG AACAGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAN	1020
	CCCNCCCCCC GNCCCCG	1036

45

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 908 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

	TTAAACAAAT GGAATCATGC AATATGTGAC CTTTTGGCTC TGGCTTATTT TATTTAGCAT	60
	AATGTTTTTG AGGTTCATCC AAGCTGTAGC ATGTATCAGC ACCTCATTTC TTTTCTGGC	120
60	TGAATATTAT TCCATTATAT GGATTTACCA CAATTCATT ACCTATTTCAT CTTTTGTTTC	180

	TGCTGTCTGG CTATTGTGAA TAATGCCTCG ATAAACATTC ATATACAAGT TTCTATGTGG	240
5	CTTTATGTT TCATTTCTCT TGGCTATCTA CATGGGAGTA GAATTCTAGG TCATAATATA	300
	ATTTTATGTT TAACTCTCA AAGAACGCC AAAAGGTTTT TCATAGTGGC TGCATCATTT	360
	ACATTCCCAC CGGCAATGTA CAAGGATTTC TATTTTCCA TATCCTTGCA CTTACCAACA	420
10	CTTCCTTTTK GTWATWATTT TGTTTTTCA TTATTGCCAC CCTAGTGGAT GTGAAATGGC	480
	ATCTTATTTGT TTTGATTTCG ATTTCTCTAA TGACAAATGA TATCATACTT TTTTTATGTG	540
15	CTTACGGATC AAAGGTATTT CCTTGGAGAA ATGTCCCTTC AAGTCCTTG CCATTTCAA	600
	ATTTGGITAT TTGTCTTTA TTATTCAGTT TTAAGAAATT CTGGCCAGGC GCAGTGGCTC	660
	ACCTGTAATC MTAGCACTTT GGGAGGCCAA GGCAGGGCAGA TCACTTGAGK TCAGGACTTC	720
20	GAGACCAGCC TGGCCAACAT GGTGAAACCC CATCTTACTA AAAATACAAA AATTAGCTGG	780
	GCGTGGTGGC AGGTGCATGT AATCNTATCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTT	840
	GAACCCAGGA GGCGGAGGCT GCAGTGAGCC AAGATCACGC CATTGCACTC TAGCCTGGGT	900
25	GACACAGA	908

30

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 655 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

	TGCACTGGTT CCTCTCCCC AGCAAATACT GCCTCTTGT TTTCTCTGA TGTGGCAGGT	60
35	GACTACAAA TCCGCCTTGG TATTCTCTAA ATGCATATAT ATTCTTTCT TGTCAGCTCC	120
	CTCTCTTCCT AGATTAAGAAA ACTGCCTCAT TTTCTGCTCA CTGGATGTGC AGTCCCAGCT	180
	TGTCTTCCCT CCCTCCCCCCC CTGTTGCAGG TGTCTTTTT TTTTTCTTC TCTCCCCACT	240
50	GGGCAGCAAA AGTGTGTCGA CAGTGGAAAW TTAGGCATCC TCAAGTTTCY TCCCAGCTTC	300
	TGCTGTGTTT TCCTAGAGTA AATTGCCAAT TTCTGTTTT ACAGGAAATC CTTTTTTAAA	360
	AATGGAATCA GTGTGGTCCC CATCTACTCT GCAAAATG CATTCTCTC TATTTCTAAA	420
55	TGAGATTTGT TCAAGTTCA AAACCACGTG AAATAATAAA TGTATAGTAG TTTCTTTTC	480
	CTTGGGCATT GCTWGATATG TGAAATGGGT TTATGAAAAA TAATAAAATC ATAACGCTAT	540
60	TTGTTTGACT TTCAATTCTA TGGGAATTCT TCTCAGCTAA ACTCTAAATG GTGATTARGC	600

AAAAAAAAA AAAAAAAACY GRAGGGGGC CCGGTACCAA TTGCCCCAT AATGA 655

5

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1102 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

15	TTTTTTTTTTT ACCATTTAAA ATAAAATGAA AGTGACCTTC TGTGTATAAA AATCTTTGTC	60
20	TGCATCTCTG CTTATTCCT TAGAAGAGAT TCCAAGAACG GGTGAGTGAT TTCAACGGCAG	120
25	CAGAGGGTTG GGACATATTA CGGGCGCGGA TCCCTCTTGG AGTGAGATGA CTCTCCGGAG	180
30	AGATTTAGTC GTCACCCCTCG CGTGTGAGGC TGGGTACAC CCCAGGGATG TGTCTATCAA	240
35	GATGGAAGAT CTTTACACG CTCTTGATTT TGTTTGSCTY TTTTCTATT ACTAGTGAGA	300
40	AKGAAACTTT TTATATGATT ATTATCCATC ATAATCCAAC ACAAAATTACT GCTTCATGTT	360
45	CTTTTACTTT CCTGTGAAGG TTTAGTGCC TTTTAAAAAT TGCTATATAT TAAGCTTGTG	420
50	AATACTTCCA TGCTGTATTT GTGGSCATCA RTTTCCCCGG GNACAGGCNT GCACATTTG	480
55	CCTTCACACG CTGGGTGGTT TTTCATTTTC AMTTCTATTT CTCGTTCTTC TATCGTTTTA	540
60	TGTTCAGACG GGTTTCTCCG TGTAGAAAGC AGTTTATGAA GATTTACTTT CGACAGTCCT	600
65	CTCTCTACTT TCTACAGTGA ATTCTCTGAT GTGTCTGGGA GTTGGGGGT CTGGGTAAGA	660
70	RTCCCTCTCT CACCTTATTTC TCTATTACGA TCCACAGCCT CATGCTTTAT GARATTGGTG	720
75	GCCGGGARCG GGGGAGATTT GCGGATCCCC CAAGCCAGAC TTTATCCCC TATCCCTGCC	780
80	TCTGGATCCC ACGTACAGGC CTGGGAACTC CCTGTGGGTA GGGCCAATG GTCTCGCACT	840
85	CTCACCTGTA CCCCAGGGCT GGACAGGGAT GGTCAAGGAG AGAGGCTGCC CAAGCGCATC	900
90	CYTCTGGTGT CCCCCTGACA CGCCCTCCAAA GTGAGCAGGT AGGTTCAAC AGCCCCACGT	960
95	TGCAGGTGGG AGATGAAGCT CAGGGTGGAG ACCAGTATCT CACAGTTCTC TTTGCATGGC	1020
100	CGGGTACTTG TTAGTCAACT GATCAAGTGA AAATTCTAGC CCCAGAGGCA GGAGAATCCG	1080
105	GAACAAAATT AAACCAGCCA GG	1102

55

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1533 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGCACGAGCC GNCACGGCA GCGCCCCATA GOGCCAGGGA CCCCCCTGGCA GCGGGAGCCG	60
10	CGGGTCGAGG TTATGGATCC AGCGGGCGGC CCCCGGGCG TGCTCCCGCG GCCCTGCCGG	120
	TGNCTGGTGC TGCTGAACCC GCGCGGCCGGC AAGGGCAAGG CCTTGCAGCT CTTCGGAGT	180
	CACGTGCAGC CCCTTTGCG TGAGGCTGAA ATCTCCTCA CGCTGATGCT CACTGAGCGG	240
15	CGGAACCACG CGCGGARCT GGTGCGGTCG GAGGAGCTGG GCGGCTGGRA CGCTCTGGTG	300
	GTCATGTYTG GAGACGGGCT GATGCACGAG GTGGTGAACG GGCTTCATGG AGCGGCCTGA	360
20	CTGGGAGACC GCCATCCAGA AGCCCCCTGTG TAGCCTCCCA GCAGGCTCTG GCAACGCSCT	420
	GGCAGCTTCC TTRAACCATT ATGCTGGCTA TRAGCAGGTC ACCAATGAAG ACCTCCTGAC	480
	CAACTGCACG CTATTGCTGT GCGCCGGCT GCTGTCACCC ATGAACCTGC TGTCTCTGCA	540
25	CACGGCTTCG GGGCTGCGCC TCTTCTCTGT GCTCAGCCTG GCCTGGGCT TCATTGCTGA	600
	TGTGGACCTA GAGAGTGAGA AGTATCGCGC TCTGGGGAG ATGGCCTCA CTCTGGGCAC	660
30	CTTCCCTGCGT CTGGCAGCCC TGGCACCTA CGCGGGCGA CTGGCCTACC TCCCTGTAGG	720
	AAGAGTGGGT TCCAAGACAC CTGCCTCCCC CGTTGTGGTC CAGCAGGGCC CGGTAGATGC	780
	ACACCTTGTCG CCACGTGGAGG AGCCAGTGCC CTCTCACTGG ACAGTGGTGC CCGACGAGGA	840
35	CTTTGTGCTA GTCCCTGGCAC TGCTGCACTC GCACCTGGGC AGTGAGATGT TTGCTGCACC	900
	CATGGGCCGC TGTGCAGCTG GCGTCATGCA TCTGTTCTAC GTGGGGCGG GAGTGTCTCG	960
40	TGCCATGCTG CTGCGCCTCT TCCCTGGCCAT GGAGAAGGGC AGGCATATGG AGTATGAATG	1020
	CCCCCTACTTG GTATATGTGC CCGTGGTCGC CTTCCGCTTG GAGCCCAAGG ATGGGAAAGG	1080
	TGTGTTTGCA GTGGATGGGG AATTGATGGT TAGCGAGGCC GTGCAGGGCC AGGTGCACCC	1140
45	AAACTACTTC TGGATGGTCA GCGGTTGGGT GGAGCCCCCG CCCAGCTGGA AGCCCCAGCA	1200
	GATGCCACCG CCAGAACAGC CCTTATGACC CCTGGGCCGC GCTGTGCCTT AGTGTCTACT	1260
50	TGCAGGACCC TCCCTCCCTC CCTAGGGCTG CAGGGCCTGT CCACAGCTCC TGTGGGGGTG	1320
	GAGGAGACTC CTCTGGAGAA CGGTGAGAAG GTGGAGGCTA TGCTTGGGG GGACAGGCCA	1380
	GAATGAAGTC CTGGGTCAAGG AGCCCAGCTG GCTGGGCCCA GCTGCCTATG TAAGGCCTTC	1440
55	TAGTTTGTTC TGAGACCCCC ACCCCACGAA CCAAATCCAA ATAAAGTGAC ATTCCCCAAA	1500
	AAAAAAAAA AAAAAAAA ANCCCGNGGG GGG	1533

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

ATCCTCTGGA ATCTAGGTGG AAGCCACCAA GCCTTCTTC	CACTTGCGTT CTGAGCATCT	60
GCAGACTTAA CCCCATGTGG CAATCACCAA GGCTTATGGC	TTGTGTCCTC CAGAAGCTGTG	120
15 GCCAGAGCTG TACCTGGGCC CCTTTGAGCT GAGGCTGAAG	CCAGAGTCTG AAGCTCAGCA	180
GGGCAGTARG GCCCTGGGCC TGGCCCTGA AACCAATTCTT	TTCCTCTAAG CCTCTGGGCC	240
20 TTTGATGGGA RGGGCTGTCC TCAAGATTT TGAAATGCCT	TTGGAGGGTT TTGCCCCPTGT	300
CTTGGATAATT GGCTTCCCTT TAGTTATGCT CATCTCTCTA	GCAAGTGAAT GTTTCACAAC	360
25 CTGCTTGGAT TCTTCTCTA CCACAGARCC AGGCTGCAA	TTTACAAAC TTTTACACTC	420
TGTTTCCCTT TTAAATATAA ATTCAATGT TAAGTCACCTT	CTTGTCTCCC ATATCTGATT	480
TAGGTGTCTG GAAGTACCCA AGTCACCTCT TGAATGCTTT	GCTGCTTAGA AATTCCTCT	540
30 ACTAGGTAGC CTGGGTCTAC ACACTTAAGT TCAA		575

35 (2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 639 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

45 TCTTTTCATC TTAAGCACCA CCCGACAGGG CAGGTACTAT	TACCATCTCC GTTTGACAGA	60
TNAGGAACCT GGCACAGGAA GCATTTAAGT GGATTCCTCA	GGATCGCCCC ACTGTCAGGA	120
50 GCAGANTCAG AATGGGCCTC AGCATCAGGC TCCCAATCCT	GGCTTCTAAC TGCTGCGCTC	180
TGCCCTTCYC TCWCCCCACC TCCCCACTCC AGTGCCTTIG	GTCATGCCAC TGCAGCTTIC	240
AGGCCAATAC TGGATTAGCC TCTTAGTGTG	CTTGTCCCTG CAGCCATTTC CCCAGGCAGC	300
55 AATTCCATGT GCCCTCACTG ATGTAGGTGG CTCCTGIGTC	ATTTGTCACA TCCTATTGAA	360
TTGTTTATGC ATCTTGTCA CACTCACAGC ACCCTCCCTC	TCACACGTCC TCCCTATAAA	420
60 AATGTCCCTC AGTGTCTGCT ATGAGCCAGG TGCAGACTTA	AGTGACAGGG CTGCTACGGG	480

AAATAAAAAA TTAACAAGGA GCACCTGCCT CTTAATGCAC AGTAACAAAC TATGTTAAGT	540
GTCAGGAAGG AAAGGTTAAG GATGCCAGGA AGGCTTTAA TAAATAACCT GACTTAGATG	600
5 GGCAGGTGGT GCTGARGATT AAGAACGTGT TCTTCTCGA	639

10 (2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 744 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

20 GAATTCCGCA CGAGAGTGGC TGGAGTCTGG CTGCAGAGGG AAGACATCAG CAGGGAGGGA	60
GCCAGGGCCT GTCACATCTT TCCTCTGGCC ATTGTCTGG TCTTTGTAAG CCCAGAAATCT	120
25 CCCCTTCCCT GAAGGGAGGC CAGCACCCCA GGAGGGCAGC AGGTGTGCTG TGAGGGTTGG	180
AGTAGTGTGA GAGGTCAGGG TACACTAGAA TGGCCATGGA CACCATGTGG GGGTGCTCTG	240
30 GGCTGGGCCA CAGAACAGTG TCCTTCCTGC TGCTCCCTCCC CTGCAGCTTC CCCCCGACCTT	300
GTNGTTTATT TGGTTTGATA CCAATCAGCA GACCCCTGCAA GGTGGAAGCT CCCAGGCTCT	360
CAGTCCCCACS ACTCTCATGT GCCAGTCACC CNTACTGTAA CTGCCCAATG AGTACTTCTT	420
35 GCCCACTGCC AAGATAGAGC CAGTTTACCA AGACAGGGGA ATTGCAGTAG AGAAAGAGIT	480
GAATATACAT AGAGCCAGCT AAATGGGAGA GTGGAGTTTT CTTATTACTT AAATCAGCCT	540
CCCYTAAAAT TCAGAGGTGA GAATTTTCA AGGACAGTTT GGTTGGSCAGG CCTAGGGAAT	600
40 GGATGCTGCT GATTGGCTAG GGATGCAATC ATAGGGGTGT AGAAAAGIWC CTTGTGCACT	660
GAGTCCACTT TTGGTGAGAG CTACCAAGGA GCTGCTGGTC TGCTGGTCCC GGTAGAGCCA	720
TCTGGTGTCA GGAATGCAAAGTG	744

45

50 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

60 GCAGGGGAAT TCGGCCACGG AGGGGTTCA ACAGGGCCCG TGGGGTGAGG TGCARACACA	60
---------------------------------------------------------------------	----

AAGCCCATAA GTGCTGGCCT GTTGGGACAA ATGAGAGAAA TCCCATAGGG TGGTGATGAC	120
ACGCCAYTCA GCCATCYTAY TCCTGGGAA AATGAAACTT GTGCTCTAT CAAATGCTCA	180
5 GTTGTAAAAC TGGAAAAAAA TTTTAGAAGA CATCTTGTC AGCATCTGTG TTTATGTCTA	240
TAAAATGTAG AAAACTAAAG CACAGAGATG TTAAATGTTT TGICCAAGGT CCAACAGCTG	300
10 GTTAGCARGC TTGGTCTGGT GACCTTTCTA CTGAACCACA GTGCCGCTGG GGGAAAGTCCT	360
CAGCACAGAT GGCTGCTGCT ATAGCTGGGG TATGGCAGT ATTAGTAGTT AACCAAGTCAA	420
CCCAAGTTCC CATACTCTAG GTTCTGCTTC AGCTGGAGGT TAGGGAAAAA CACAAGAAAA	480
15 TCCCTTACCA CTCTACCAGT GCTGGGGAT GTACTAAGAG ATCCCC	526

20 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

30 GGCACAGGGC AGGAGAGACT TGGTCCATGG GGAGAACCT GCAGTATAGA TGGGACCTCC	60
AGGAGCCCAA GTAGCATAGA CCCCTGCTGAT CCCGGGCCAT TGAGCCAGAG GATTTGGCT	120
35 GAATGTCCCC AGAGACAAAA GGGAAAGGTA GATCCTTTCC CTTAAAGATG AAAGCCATCG	180
CCCGGGCTTG CTTATTGCTC TCTCTCCTGG TCCCTCCACA TGTIGTTCT GAACATTTGT	240
TCTGGCATCA CAATCCCCGT CATCCTGTCA TCTGGCCCTT CCCACCTTTC CACCTTATCT	300
40 CTTGCAGTGT CTCCGGTCTG ACCTGGCACC TGGGTGAARG CTTGCTCTTG CTGGTGCCCA	360
TAGCCCCCAG TGTATGGTCT TGAMCTCCCC AGCCATATGG ARACCCACCT CAGGAGGCC	420
45 CCTCGA	426

50 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 844 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60 GGCACAGCGG CACGAGATAG GAAGCTTGGC AGGGCCAGCT CCCCCAGTGC GCATTGCCCT	60
----------------------------------------------------------------------	----

	GTAACACTGAG CGCCTGGGAG TGGGGAGAGG CTTGGAAATG GAGCACGGTG GTGGACCTCG	120
	TCTTCTCCTG CTCATCCCAG GCCTCCTCCA TAACACCTAC CTAGCACGGC CTGGGGACTT	180
5	CCCAGCCCAA GGAACAACGT AGAATACTGA GTGCCAGGGT AGCCCTAGCC CCATTCACA	240
	CCTGGGCAAA GTGAGGTACAC TGGATTCAAA CACTCAGATT TAAACCTCCT CTGTGTCTGC	300
10	AGCACCTGTA TATAACTGCC AGCCTCTGCT GCCCCTCTCC AAAAAGTCAC TGCCCTTGTC	360
	TTTGGCACCT GTCTCTGTCC TCCCCATTCT CTGCTCCTCC TTTCTCCAAC TCAGANTCAC	420
	CCTGTTAGTT CAGCAAATGT TCATCGAGCT CCATAATGTA GCAGGACAGG NCTGTCTAAC	480
15	AGATTCTGGN CTTGCAAGGG TGAGACAAGT ACTCTCCATC TTTCTCTCAT CTTCACAGAT	540
	GGTCTGCTCA ACAACTTTGC ACTGAATTGT AAATAATTGA TACTGCATAA AACATTGATG	600
20	TTCTTTAAGG GTAGTCCAGC AAGGTGCCAA GTCTTATAAT GATAACTGCT CAAGGATCTC	660
	TCAGTGAAGC ATTTGGGGST GCTAGCTCTG CCTATGGGTG AGGTCAAGCTA TCTCACGCCA	720
	TCTACTTCCA CNTGCCCCCC CATGCCAGGC TCACCCCTGAG CTGAGATGCC TGAGCAGGTG	780
25	GCAGAAAAGGA GCCACCTGGT TTATGCTTCG GGACCACAAA CTCCCTATC CAGANGACAG	840
	TTTT	844

30

(2) INFORMATION FOR SEQ ID NO: 97:

35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1985 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDENESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:	
	AGCCCTGCTG AAGTACAGGT TCTTCTATCA GTTCTCTGTTG GGCAATGAAC GAGCAACAGC	60
45	AAAGGAGATC AGGGATGAAT ATGTGGAGAC GCTGAGCAAG ATTTACCTGT CTTACTACCG	120
	CTCTTACCTG GGGGGCTCA TGAAGGTGCA GTATGAGGAA GTGCCGTGAGA AAGATGATCT	180
	AATGGGTGTG GAAGATACAG CAAAGAAAAGG ATTCTYCTCA AAGCCATCGC TCCGCAGCAG	240
50	GAACACCATT TTCACCCCTAG GAACCCGGGG CTCTGTCTAC TCCCCACTG AACTTGAGGC	300
	CCCCATCCTG GTGCCTCACA CAGGCCAGCG GNAGAGCAGA CGTATCCATT TGAGGCCCTC	360
55	TTCCCGCAGCC AGCACTACGS CCTCCTAGAC AATTCTGAGC GCGAATACCT TTTCATCTGT	420
	GAATTTTTTG TTGTGTCTGG CCCAGYTGCA CACGACCTGT TCCATGCTGT CATGGCCCGT	480
	ACACTCAGCA TGACCCCTGAA ACACCTGGAT TCTTATCTAG CTGACTGCTA CGATGCCATT	540
60	GCTGTTTTTC TCTGTATCCA CATTGTTCTC CGGTTCCGTA ACATTGCAGC AAAGAGGGAT	600

	GTTCCTGCC	TGGACAGGTA	CTGGGGAACA	GGTGCTGCC	TTGCTATGGC	CACGGTTGA	660
5	ACTGATCCTG	GAGATGAATG	TTCAGAGCGT	CCGAAGCACT	GACCCCCAGC	GCCTAGGGGG	720
	GTGGATACT	CGGCCCCACT	ATATCACACG	CGCTATGCA	GAGTTCTCCT	CCGCTCTTGT	780
	CAGTATCAAC	CAGACAATT	CTAATGAACG	GACCATGCAA	TTGCTGGGAC	AGCTGCAGGT	840
10	GGAGGTGGAG	AATTTGTCC	TCCGAGTGGC	AGCTGAGTTC	TCCCTAACGA	AGGAGCAGCT	900
	TGTGTTCTG	ATCAACAAC	ATGACATGAT	GCTGGGTGTG	CTGATGGAGC	GGGCTGCAGA	960
15	TGACAGAAA	GAGGTTGAGA	GCTTCCAGCA	GCTGCTCAAT	GCTCGGACAC	AGGAATTCAT	1020
	TGAAGAGTTG	CTGCTCCCC	CTTTTGGGGG	TTTAGTGGCA	TTTGTGAAGG	AGGCTGAGGC	1080
	TTTGATTGAG	CGTGGACAGG	CTGACCGACT	TCGAGGGAA	GAAGCCCCGG	TAACTCAGCT	1140
20	GATCCGTGGC	TTTGGTAGTT	CCTGGAAATC	ATCAGTGGAA	TCTCTGAGTC	AGGATGTAAT	1200
	GCGGAGTTTC	ACCAACTTCA	GAAATGGCAC	CAGTATCATT	CAGGGAGCGC	TGACCCAGCT	1260
25	GATCCAGCTC	TATCATCGCT	TCCACCGGGT	GCTGTCCCAG	CCGCAGCTCC	GAGCCCTCCC	1320
	TGCCCGGGCT	GAGCTCATCA	ACATTACCCA	CCTTATGGTG	GACCTCAAGA	AGCATAAGCC	1380
	CAACTTCTGA	TGTGCCAGAA	ACCGCCCTGA	GATCTGCCGG	TCATCTCCAT	GGACTTCTGC	1440
30	ACCCCATTCC	ATACCCTTCT	TCACCTGGGG	TACCCCTTCC	AGTTTTCCCC	TTGCTTCCCA	1500
	GGCCCTTGAC	ATGGCTTAC	TGCCCTCACT	CCCAGCACCT	TGCCCAACAG	GATAAGCTGG	1560
35	ATCCCCTTGG	CCTTCTGAAT	ATCCCAGTGT	CTTCAGGTTT	CCCAAGACCA	CTTCCCTGTG	1620
	GGCTTCCAAA	ATGGCCTTTA	TCATTTCTCC	AGTCTGTAC	CCTCCCTTCC	TGCTCCCATA	1680
	CACCCAAGGC	TTGTTTCTTC	CCCTGTAAAAA	ACCACTGCC	CAATCTCTGG	TTCACTCAAC	1740
40	TAGTCACCAT	GTCCTGAGGC	ATGAAGCCTC	CTCAGCTCTT	GGAAATTGCTG	GCAAGGGGTG	1800
	ACTGCCTCTG	AGTCATTGTG	TTTTTCAAAG	TGATTTCTTT	TCTGTAGCTT	TTTGACCTAA	1860
45	GATCTCAGCA	ATTTGAACAC	TAACCTCTCC	CCTCCCTGGCT	CAAGAATTAC	TCCGAAGTCA	1920
	GTCTGCAGAA	AATAAATATT	TAGTATGACA	TGAAAAAAA	AAAAAAA	AAAAAAA	1980
	AAAAAA						1985
50							

(2) INFORMATION FOR SEQ ID NO: 98:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1416 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

	ATATGAAGGG AAAGAATTIG ATTATGTTTT CTCAATTGAT GTCAATGAAG GTGGACCATIC	60
5	ATATAAATTG CCATATAATA CCAGTGATGA CCCTTGGITA ACTGCATACA ACTTCCTTACA	120
	GAAGAAATGAT TTGAATCCTA TGTTTCTGGA TCAAGTAGCT AAATTTATTA TTGATAACAC	180
10	AAAAGGTCAA ATGTTGGGAC TTGGGAATCC CAGCTTTCA GATCCATTAA CAGGTGGTGG	240
	TCGGTATGTT CGGGGCTCTT CGGGATCTTC TAACACACTA CCCACAGCAG ATCCCTTTAC	300
	AGGTGCTGGT CGTTATGTAC CAGGTTCTGC AAGTATGGGA ACTACCATGG CGGGAGTTGA	360
15	TCCATTTACA GGGAAATAGTG CCTACCGATC AGCTGCATCT AAAACAATGA ATATTTATTT	420
	CCCTAAAAAA GAGGCTGTCA CATTIGACCA AGCAAACCT ACACAAATAT TAGGTAAACT	480
20	GAAGGAACCTT AATGGAAC TG CACCTGAAGA GAAGAAGTTA ACTGAGGATG ACTTGATACT	540
	TCTTGAGAAG ATACTGTCTC TAATATGTAA TAGTTCTTCA GAAAAACCCA CAGTCCAGCA	600
	ACTTCAGATT TTGTTGGAAAG CTATTAAC TG TCCTGAAGAT ATTGTCTTTC CTGCACTTGA	660
25	CATTCTTCGG TTGTCAAITTA AACACCCCAG TGGAATGAG AACTCTGCA ATGAAAAGGA	720
	AGGGGCTCAG TTCAGCAGTC ATCTTATCAA TCTTCTGAAC CCTAAAGGAA AGCCAGCAA	780
30	CCAGCTGCTT CCTCTCAGGA CTTTTGCAA TTGTTTGTGTT GGCCAGGCAG GACAAAAACT	840
	CATGATGTCC CAGAGGAAT CACTGATGTC CCATGCAATA GAACTGAAAT CAGGGAGCAA	900
	TAAGAACATT CACATTGCTC TGGCTACATT GGCCCTGAAC TATTCTGTTT GTTTTCATAA	960
35	AGACCATAAC ATTGAAGGGA AAGCCCCATG TTGTCACTA ATTAGCACAA TCTTGGAAAGT	1020
	AGTACAAGAC CTAGAAGCCA CTTTTAGACT TCTTGTGGCT CTTGGAACAC TTATCAGTGA	1080
	TGATTCAAT GCTGTACAAT TAGCCAAGTC TTAGGTGTT GATTCTAAA TAAAAAAGTA	1140
40	TTCTTCAGTA TCAGAACCAAG CTAAGTAAG TGAATGCTGT AGATTTATCC TAAATTTGCT	1200
	GTAGCAGTGG CGAAGAGGGGA CGGATATTAA TAATTGATTA GTGTTTTT CCTCACATT	1260
45	GACATGACTG ATAACAGATA ATTAAAAAAA GAGAATACGG TGGATTAAGT AAAATTTTAC	1320
	ATCTTGTAAA GTGGTGGGGA GGGGAAACAG AAATAAAATT TTTGCACTGC TGAAAAAAA	1380
	AAAAAAAAAAA AAAAGGAAAC TCGAGGGGGG GCCCGG	1416
50		

(2) INFORMATION FOR SEQ ID NO: 99:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

	NTCTACCCCTA ATCAAGATGG GGACATACTT CGCGACCAGG TTCTTCATGA ACATATCCAG	60
5	AGATTGTCTA AAGTAGTGAC TGCAAATCAC AGAGCTCTTC AGATACCAGA GGTTTATCTT	120
	CGAGAACGCAC CATGGCCATC TGCACAATCA GAAATCAGGA CAATAAGTGC TTATAAAAACC	180
10	CCCCGGGACA AAGTGCAGTG CATCCTGAGA ATGIGCTCTA CGATTATGAA CCTCCTGAGC	240
	CTGGCCAATG AGGACTCTGT CCCTGGAGCG GATGACTTTG TTCCCTGTGTT GGTGTTGTG	300
15	TIGATAAAAGG CAAATCCACC CTGTTTGCTG TCTACTGTGC AGTATATCAG TAGCTTTAT	360
	GCTAGCTGTC TGTCTGGAGA GGAGTCCTAT TGGTGGATGC AGTTCACAGC AGCAGTAGAA	420
	TTCATTAAGG CCGAAAGTGA CCAAGACCAA GGCCCACCAA GGCAGCAGAC	480
20	TGTTAACAGA ACAAAACAGAT CTCTGAGAAG GTGCATCAGC TGCTTGAAG GCTGAAGATT	540
	GTGTTGTATG ATACTGCACA GCATCAGGCA TTTTAAAGCA GATCTTTACT AAACAGGTTA	600
25	ATGAGCTAAC AAGCAGGTC TCTCGTCTTT GGGCTCTTC CTTCTGAGT TGCATATTCT	660
	ATTTCTTGT CCCAAGTAG AGACTAGTAC TACAAAAAGG GACCACATTT TTCAAGTATT	720
	TCTAAGTATA AAAAACAAAA CAAAATCTC TTAGGAAATG TCTAGACCTC CATTCTTGGG	780
30	TTCCCTTTCT TTCCCTTTAT TTTAAAAAG AACAGTACCC CTCTTTAAG ATGCTGTCTT	840
	ACATTAATGA GCATCTAACG GAAAGAAGGT ATGAGTIGCA CTGAGGATTA GAATAGTGGT	900
35	GGGTTAGTGG CATTATCTAT AAATACACTC ACCTAAATTG AAAGCTAAGA AGGAAATGTA	960
	AATATAATAT ATATTTATAT TTGATGTAAT ATGGACATCT GCAGATTCTA ATAAACAAGG	1020
	ACTATTGCTG ATAGTAGGCT GTGACATACT GTCTTGTGAA ATGGTTTCCT TGACAAAATT	1080
40	TAAGCTGAGC TTAAAAGCAA AAAAACAAAA AGTACACAGA AATATTATT AAAATGTAAT	1140
	ACAGTTTATT GAACTTTCTA GGTATGGAGT TTGATGGACA GGGCTGCCY TAATGAGTGT	1200
45	GAAGGTCACT AAGTCACITTA GACATCTCAC CGTGGAAAGTT TGTGAGCCTG CATTAGGAGA	1260
	TAGACTGATT ACCATACATG ACATAAAAAG GAACAGTGGG TAGCTCATAAC TTTATGGTGG	1320
	TTCTTCTCCT CCGAAATAAT ATACTGCAGA AATCCAGAC AGAGCTCCTT ACAAAACCTT	1380
50	AATTGTAATA TATTTTGAT GATTATTCAAC ATTGAATGCA CAGACCAAGA ATTCAAGTGA	1440
	TGTCATTTTT TAAAAAAACTA ATTGTATTG TCTGCTCTAG TGATACAAGT TTTACTAGTG	1500
55	ATAAACTATT TTAATCAACC ATACTATTCT TATGGAAAAA AATATCTATT TTGGCAGGTT	1560
	TCTGTGCCCTT TATTCCTC TTCTGAAAAA AAGTCTGTGT TTTCATAGTT TGGTTTGCAT	1620
	TGTATATCAA TAATTAATCA GGAATGGGTT TTGGTGCCTG AAAAATTGGC CATGGAGGCA	1680
60	CACCAAAGCT TCAAGCACAA GTCTTGTACA TGGGCCATCA CTGCTGGTT TCACTTCGTG	1740

TGTTTCCTAA ACACATTTAG CTGCCTTTT AACAAACTCA GCCCCATACT TGAGTCCCTT 1800
 5 GTTGTGGGA GCATTTCCAG GCATCTTTA AGGAACTGT GACAAACAGC CTCGGCAGA 1860
 TGAACACCGA GGCTCTCTGT TGTCTGTCTC TGAGATCTTT GTGTCTGGGA ATGCCTAAAG 1920
 NTTTGNTTT TTTTT 1935

10

(2) INFORMATION FOR SEQ ID NO: 100:

15

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 599 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GAATTGGCA CGAGCGTCCA CGCAGCCGCC GGCGGCCCAG CACCCAGGGC CCTGGCATGCC 60
 25 AGGTGTTGG AGGTGGCAGC GAGACATGCA CCCGGCCGG AAGCTCCTCA GCCTCCCTTT 120
 CCTCATCCTG ATGGGCACTG AACTCACTCA AGACTCCGCT GCCCCCCGACT CCCTGCTGAG 180
 30 AAGTTCAAAG GGCAGCACGA GGGGGTCTTT GGCTGCTATT GTCATCTGGA GGGGAAGAG 240
 TGAGAGCCGG ATAGCCAAGA CCCCAGGCAT TTTCAGAGGT GGGGGGACCT TAGTCCTACC 300
 CCCAACACAC ACCCCTGAGT GGCTCATCCT CCCTTGGC ATAACGCTGC CCTTGGGGC 360
 35 TCCAGAAACA GGCGGTGGGG ATTGTGCCGC TGAGACCTGG AAGGGCAGCC AGCGTGCCGG 420
 CCAGCTGTGT GCATTGCTGG CTTAATATGC AGGGCTTGGG GGGCTGTGGC CACATGCCCG 480
 40 GCAGGAGGTG AGTGAGGGAGC CCTGTGGCGT GCTGGTGTGG GGATCGTGGG CATTTCAAAC 540
 GGGCTTGTGG TACCCCTGAAC AATGTATCAA TAGAGAAAAA AAAAAAAAAA AAAACTCGA 599

45

(2) INFORMATION FOR SEQ ID NO: 101:

50

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 784 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

GAATTGGCA CAGAAAAAAA AGAGAGACTG GGTCTTACTG TGTTGCCAG ACTTGTCTTG 60
 AACTCCTGCC TCAGCCTCTC AAGTACTTGG GATTATAGGC CAAGAAGCCA CCATGCCTAG 120
 60 CTTCTTCTTG TCATTGATCC AGACTAATAC TCTGGGTCA GCCTCATTTC TTCTCTTCT 180

	CACTTTGCAC ATCCACTTGT CACCAAATCK RGTCATTCT GCATCCTAAG TAAGTCCTTT	240
5	GATTCCCTCCA GTGGMTCATT AGTAATGTCT CAARTGTAAT TTTTTCTAGT AGTTTCAGC	300
	CIGTCTTCC KGCCCTCAGT CTTAACTTCT CCAGTACATA KGCCACATTG TTGTCAGCAK	360
	GATCAWATT TATTTAAAAA TACTTTACAW AKGTTATKG CCAAATATTA GRAAATACAG	420
10	ATTCAATGGAA AGAAAAATCA CTGTCCTCAAG GAGGTCACTG GCATGGTGAG GTTAAGGGT	480
	GATTTTAATT TTTAAAAATG TATATTTTTT CCTGTGAGA GTAGTAACAC CCTTGAAAAC	540
15	ACAWTCCCTT GTAAAGTCTC TAATTCTGTA CTCCGCATCT AGSTGRTCTC TTCTTCTCA	600
	GATATTTAC AATTTCAATT ATCACCACCT TTCTCTAGCC TTTACCCGTC TCTTCAATAT	660
	TWACATATGC AGAAGTTTCT CCTAACAAAC ACCTGCCCT GCCTCAGTTC TGCTACCACC	720
20	CTGTTGCTTT CTPTCCCTTC ACAATCAAAT TTAAGAGTGT CAAAAAAA AAAAAAAAAC	780
	TCGA	784

25

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:

30	(A) LENGTH: 1035 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

	AGAGGCCTGG CTGCGTTGCC CTATCTCCGT CTCGCCACC CACTTAGCGT TTTAGGCATC	60
40	AATTACCAGC AGTTTCTCCG CCACTATCTG GAAAATTACC CGATTGCTCC CGGCAGAATA	120
	CAAGAGCTTG AAGAACGCCG CAGTTGCGTG GAAGCCTGCA GAGCAAGGGA ACCAGCGTT	180
	GATGCCGAAT ATCAGCGAAA TCCTCACAGG GTGGACCTCG ATATTTAAC CTTTACGATA	240
45	GCTCTGACTG CCTCTGAAGT TATCAACCCT CTGATAGAAG AACTTGGTTG CGATAAGTT	300
	ATCAATAGAG AATAGTTAGG TGGTGACACT ACTTCAGAG AACCTCTGCA TTCCAGTCAT	360
50	ACCAATCCTG CAACTTGATT TTCAAGTC AAGAGTATAT CGCGATAAGA CAGTGCACAG	420
	GTGGAGGGGA AAAAAAGGGG GAGGGGAAG CTTATCTTGA AAAAGCATCA CAGAAGTAGA	480
	AAAAAAATGTC GAAAGCATTA TAACTGTAAC GTTCTTTGAG TTTGTGATTG ATCCACATTT	540
55	TTCCCCCTGC ATTATGGAAA ATGTCCTCTCA GCATTGCTTT ATTACAAAGT AAAGGATGGT	600
	TTTATAAAAT TGAGACTGAT GAAACATCAA TACTAGAGCC CATGAGGATG AAAGAAATTA	660
60	TCAAATAGTG CTGAACAGAA TAAGATGTTA ACGCTGAGTT ATTAGGACTG GAAGGCTATG	720

	AAAAGAACTT GAAATTGTCG GAATATGTGC TCTCTTCATG TCATATTCAA TAGAAGTTTC	780
	TAGTTTAAGA TTGATTTTGT GTTTCTTAG GCATTTCAAG TGACAAGCAA AGTAAATGTA	840
5	TATATTATGT GATAAACAT GTTTCAAGA ACGTCAAATT TCTGGACTTT TTTCTTCAA	900
	TTTTTAATTT TTAAAGTTTT TTTGGTATTA AAAAATCYAT TCACAAGCCA AAAAATWTWT	960
10	WAAATWIWCM CGGAAAAGCC AAAAAAAA AAAAMMAGGG GGGGCCGGC CCCATCCCC	1020
	CAAGGGGGTC CNGNT	1035

15

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2218 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:
 AGGTATTAGG CCCTTTGTG GGAGCCCCAT GTTTGTTTT TCTGAGTTGG TGGGGAGGGA 60

SGGAGGGGGA GGGCTGAATT GTTTGCAGA GGAAGATGGC ATCTGTGCTT TAAATTTCTC 120
 30 ATTACTGGGT TAGAAAACAA AGAGGGAKTG CCCTGCACAT TTTCTTTGT GCTTTAAAT 180

GTTTCTTAAG TTGGAACAGG TTTCCTCGGG CCTGTTTGA CTGATTGCTG GAGTGCATT 240
 35 GATAGTTAAA AATTACTAAT TGGTTTTATT TCCCTTCACA CTCTGCCTCC CCACTCTCC 300

CCCCGTTACT GAAAATAAC CATTAGTGT TCAGGCTAGA AATTGAATTG CTGAGTTTG 360
 TGATCCTTT AAATTAAAAA CCACAAGTGT TTATTGTAGT GGTTAAACTG TAGCATCTCA 420

40 GCATCTGGGT GGAAGCTGCC TATATTTCTT CCCAGTTAA CTGGGGACCA TCTGTGAAAT 480
 TAATTTCCA TCCAGACAGC TGCTGTGAGC AAATGAACAT AAATGCTCGC TGGAAATTAA 540

45 CTAACCAGTT TTTATATTGA CCTGCAGTGT AAAAAGCACA TTTAATTATA AACAAATATAT 600
 TCAAAATGGG CAAATTATAT TTTCAAATGC AGTGTAGAGC TAGATTAAAA GCAACTCTTT 660

GCCACCTACT CTGCCCTTT GGCAAAGTTA CCTGAAACAA AGAACCTTAA GGGTTTATTA 720
 50 AGAACTCTTT ATTTCTTCA TACCCCTGTC TCTGCAGTGC TTTCTAACAG CTTCTGGGTG 780

CAGATTTCTC TCGGCATCCT TTTGCACTCA GCTTATTACA GGTAGGTAGT GCTTAAGAAA 840
 55 AGTCATGGAG GACTAAAGCC TAAGTCCTTT TCACTTTTCC TCCATCTGAA GGTAGGTGAG 900

TTCATCCTCT TCATAGTAAT GCTGTTTAC CAAGACTTTA TAGCAGATGG ACCCAGAAAG 960
 AATTTCTGTC TATTGTGTT ACTACAACAG GATAGGGACA TCAGACAGCC CCAGAAACCC 1020

60 CTTCCAGATC TGATATGGGA CTATTAATT TTATGCTGTT AATTGGTATT CATTCAAAAT 1080

	GCAGTTGAAG GGGGAAGGCT CCACTGCATT CTTTGGCTAA GGCTGAATG CTTGCTCATC	1140
5	TGTAAGATCT ATACTCGAGG TTTTGTTTTC CTTTTAAAAT TCTTTAGGGA GAGAGGGATG	1200
	GTTTCTGAGG GGTTCTGAAA GTATGATTCA ATGTGCAACA TACAGGTAGG TCTTCAGCAT	1260
	AAGCTGAAAT ATATGCATGT AAAAACTTTG ACATCTTTT TTTAATTCTT CCACTTTCTT	1320
10	CTTAACCTTA CTTCTCTTT TGTCCCCCCCC CCATCTTACA GAAGTGTAGG CCAAGGGAGA	1380
	ATGGTAGGCA CAGAAGAAC ATGGCAAAC GCTCTGTGCT TTCAAACCAA AGTGTCCCC	1440
15	CCAACCCAA ATTGTCTAA GCACTGGCCA GTCCTGGTG GGCATTGTTT TCTACAACCA	1500
	AATTCTGGGT TTTTTCTTC TTCTTTAAA CATAGACGTA CCACCACAAG GGATGCCCTA	1560
	CTCTCTCGCA GCTCTTGAAA GCATCTGTTT GAGGGAAAGG TCTCTGGCA ACCAAGTGGT	1620
20	TATTTGGATT GCTTGCTTCC CTTTTTCCAC CTGGGACATT GYAATCATAA AATAACAGTA	1680
	AATTCCAAAC CTCAAAACT ATTATGGCT GAGCACAGCT GAAATCTAGC AGAGTTAAC	1740
25	TCTTCTGCCT CCATGTCGT CACITATAAT TCAGGTTCTG CTGTTGGCTT CAGAACATGA	1800
	GCAGAAGAAC CGTTTTATGC TAGTTATTGC ATTCAATGGTT GAAACTCAAC TTAGGGAAAG	1860
	GGTTCCAATG TATTAAGCAA TGGCTGCTT CTCCCCAATC CTCCCTAACCA ATTCTGGTG	1920
30	TGGACTTCTC ATCTAAAAGG TTAGTGGCTT TTGCTTGGGA TCAGTGTCT CTATTGATGT	1980
	TCTTGCTGGT CTCCAGACAC ATTCCCTGGTG CATTAAAGACT TGAAAGACTT GTAGATGTGT	2040
35	GATGTTCAAG CACAGGATGC TGAAAGCTAT GTTACTATTG TTAGTTGTA AATTGTCCCTT	2100
	TTGATACCAT CATCTTGTGTT TCTTTTGTGAA GGTATAAATA AAAACACTGT TGACAATAAA	2160
	AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAA	2218
40		

(2) INFORMATION FOR SEQ ID NO: 104:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

	CTTCACAGAC TGACAGAACG GTTTTGTGTTT GTTTTGTGTTT GTTTTGTAGA	60
55	TGGACTCTAG CTCTGTCACC CAGGCTGGAG TGCAGTGGTG CGATCTCGGC TCACIGCAAG	120
	CTCCGCCTCC CGGGTTCTCA CCATTCTCCT GCCTCAGCCT CCCGAGTAGC TGGGACTACA	180
60	GGCGCCCCACC ACCACGCCCG GCTAATTGTTT TGTATTGTTT AGTAGAGACG GGGTTTCACC	240

	ATGTTAGCCA GGATGGTCTC GATCTCCGGA CCTCGTGATC CGCCCGCYTC GGCGCTCCAA	300
	AGTGCTGGGA TTACAGGCCT GAGCCACCGT GCCTGCCCA GAATGGTTTT TAAAGCCACA	360
5	GTTGAGARGC CACCCATTGC CGGGCGCTG GACAGTGATC ATCTTGTCA TCTTGTTCAG	420
	TCCCTTCITG TGTGATTGGA ATTATTCATC CCCTTGAAA GATGAGAAGG TTGAGATGCA	480
10	AAGAGTCTAC CTTCACAAGT TCTCACTGCT GGAAAGARCT AGAACACAG TTCAAAGTTC	540
	TGGNTTCTGG ACTCTGCAGT CCAGGTYTCC CTTYTCCCAC TTGCCTACCC TCAATGCCAC	600
	ACTGTTTTTG AAGTGGCCA TAACTTGAAG GRAAAGTTA AAGACAGTTC AATTTAATCA	660
15	TCAGRATGCA TTCTTTTTT TTTCGGARAC GGARTTCAC TCTTGCTGCC CASGCTGGAG	720
	TGCAATGGTG CAATGATCTC GGCTCACTGC AACCTATGCC TCCTGGGTTTC AAGNGATTAT	780
20	CCAGCCTCAG CCTCCCGAGT AGCTGGGATT ATGGGGCCC ACCACCATGC CCAGCTAATT	840
	TTTGTATTIT TTTTTTTAGT AGAGATGGGG TTTCGCCAGG TTGGCCAGGC TGKTCTTGTG	900
	AAVTCCTGGC YTCAGGTGAT YTGCCCACYT CATCYTCCAA AAGTGCTGGG ATTACAGGCA	960
25	TGAGCCACTG CGCCTGGCYT CAGAATGCAT TCTTACACAT CTATCCTAGA CATTATAAG	1020
	CACTCTAAATG GATAACAATC CAAGAATAAA TGATTGTAAA AGATGATGCC GAAGAGTGA	1080
30	TGTCAATCTT TTTTCCTAA GAAAAAAAGT CCGCGAGTAT TAAATATTAA GATCAATGTT	1140
	TATAAAATGA TTACTTTGTA TATCTCATTA TTCTTATTAA GGAATAAAAAA CTGACCTTCT	1200
	TTAACATCATAT ACTTGTCTTT TGAAATAGC AGCTTTGTG TCATTCTCCC CACTTTATTAA	1260
35	GTTAAATTAA ATTGGAAAAAA ACCCTCAAAC TAATATTCTT GTCTGTCCA GTCTTATAAA	1320
	TAAAACTTAT AATGCATGTA AAAA AAAAAAA A	1351

40

(2) INFORMATION FOR SEQ ID NO: 105:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2066 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

	GGCACGAGGC GGCAGGAGGC CACAATCACA GCTCCGGCA TTGGGGGAAC CCGAGCCGGC	60
	TCCGCGGGGG GAATCCGTGC GGGCGCCTTC CGTCCCCGTC CCATCCTCGC CGCGCTCCAG	120
55	CACCTCTGAA GTTTTGCAGC GCCCAGAAAG GAGCCGAGGA AGGAGGGAGT GTGTGAGAGG	180
	AGGGAGCAAA AAGCTCACCC TAAAACATT ATTCAAGGA GAAAAGAAAA AGGGGGGGCG	240
60	CAAAAATGGC TGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC ATTGTGGTG	300

	GGATTCTGCT CGTGTTCAA ATCATCGCCT TTCTGGTGGG AGGCTTGATT GCTCCAGGGC	360
5	CCACAAACGGC AGTGTCTAC ATGTCGGTGA AATGTGTGGA TGCCCGTAAG AACCATCAC	420
	AGACAAAATG GTTCGTGCCT TGGGGACCCA ATCATTGTGA CAAGATCCGA GACATTGAAG	480
	AGGCAATTCC AAGGGAAATT GAAGCCAATG ACATCGTGT TTCTGTTCAC ATTCCCCTCC	540
10	CCACACATGGA GATGAGTCCT TGGTTCCAAT TCATGCTGTT TATCCTGCAG CTGGACATG	600
	CCTTCAAGCT AAACAACCAA ATCAGAGAAA ATGCAGAAGT CTCCATGGAC GTTTCCCTGG	660
15	CTTACCGTGA TGACGCATTG GCTGAGTGGG CTGAAATGGC CCATGAAAGA GTACCACGGA	720
	AACTCAAATG CACCTTCACA TCTCCCCAAGA CTCCAGAGCA TGAGGGCCGT TACTATGAAT	780
	GIGATGTCTT CCTTTTCATG GAAATTGGGT CTGTGGCCCA TAAAGTTTAC CTTTTAAACA	840
20	TCCGGCTGCC TGTGAATGAG AAGAAGAAAA TCAATGTGGG AATTGGGGAG ATAAAGGATA	900
	TCCGGGTGGT GGGGATCCAC CAAAATGGAG GCTTCACCAA GGTGTGGTTT GCCATGAAGA	960
25	CCTTCCCTAC GCCCAGCATC TTCATCATTA TGGTGTGGTA TTGGAGGAGG ATCACCATGA	1020
	TGTCCCCGACC CCCAGTGCTT CTGGAAAAAG TCATCTTGC CCTTGGGATT TCCATGACCT	1080
	TTATCAATAT CCCAGTGGAA TGGTTTCCA TCGGGTTTGA CTGGACCTGG ATGCTGCTGT	1140
30	TTGGTGACAT CGACAGGGC ATCTTCTATG CGATGCTTCT GTCCTCTGG ATCATCTCT	1200
	GTGGCGAGCA CATGATGGAT CAGCACGAGC GGAACCACAT TGCAAGGTAT TGGAAAGCAAG	1260
35	TCGGACCCAT TGCCGTGGC TCCCTCTGCC TCTTCATATT TGACATGTGT GAGAGAGGG	1320
	TACAACTCAC GAATCCCTTC TACAGTATCT GGACTACAGA CATTGGAACA GAGCTGGCA	1380
	TGGCCTTCAT CATCGTGGCT GGAATCTGCC TCTGCCTCTA CTTCCGTGTT CTATGCTTCA	1440
40	TGGTATTTCAT GGTGTTTCGG AACATCAGTG GGAAGCAGTC CAGCCTGCCA GCTATGAGCA	1500
	AAGTCCGGCG GCTACACTAT GAGGGGCTAA TTTTAGGTT CAAGTCTCTC ATGCTTATCA	1560
45	CCTTGGCTTG CGCTGCCATG ACTGTCACTCT TCTTCATCGT TAGTCAGGTA ACGGAAGGCC	1620
	ATTGGAAATG GGGCGCGTC ACAGTCCAAG TGAACAGTGC CTTTTTCACA GGCATCTATG	1680
	GGATGTGGAA TCTGTATGTC TTTGCTCTGA TGTTCCTGTA TGCACCATCC CATAAAAAC	1740
50	ATGGAGAAGA CCAGTCCAAT GGAATGCAAC TCCCATGTAA ATCGAGGGAA GATTGTGCTT	1800
	TGTTTGTTC GGAACATTAT CAAGAATTGT TCAGCGCTTC GAAATATTCC TTCATCAATG	1860
55	ACAACGCAGC TTCTGGTATT TGAGTCAACA AGGCAACACA TGTTTATCAG CTTTGCATT	1920
	GCAGTTGTCA CAGTCACATT GATTGTACTT GTATACGCAC ACAAAATACAC TCATTTAGCC	1980
	TTTATCTCAA AATGTTAAAT ATAAGGAAAA AAGCGTCAAC AATAAATATT CTTGAGTATA	2040
60	AAAAAAAAA AAAAAAAAAA AAAAAA	2066

5 (2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1705 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15	AATTCGGCAK AGGGCAGCTG TCGGCTGGAA GGAACTGGTC TGCTCACACT TGCTGGCTTG	60
	CCGCATCAGGA CTGGCTTTAT CTCCCTGACTC ACGGTGCAAA GGTGCACTCT GCGAACGTTA	120
20	AGTCCGTCCC CAGCGCTTGG AATCCTACGG CCCCCACAGC CGGATCCCCCT CAGCCTTCCA	180
	GGTCCCTAAC AC TCCCCTGGAC GCTGAACAAT GGCCCTCCATG GGGCTACAGG TAATGGGCAT	240
	CGCGCTGGCC GTCCCTGGGCT GGCTGGCCGT CATGCTGTGC TGCGCGCTGC CCATGTGGCG	300
25	CGTGACGGCC TTICATCGGCA GCAACATTGT CACCTCGCAG ACCATCTGGG AGGGCCTATG	360
	GATGAACATGC GTGGTGCAGA GCACCGGCCA GATGCAGTGC AAGGTGTACG ACTCGCTGCT	420
30	GGCACTGCCG CAGGACCTGC AGGGGGCCCG CGCCCTCGTC ATCATCAGCA TCATCGTGGC	480
	TGCTCTGGGC GTGCTGCTGT CCGTGGTGGG GGGCAAGTGT ACCAACTGCC TGGAGGATGA	540
	AAGCGCCAAG GCCAAGACCA TGATCGTGGC GGGCGTGGTG TTCCCTGTTGG CCGGCTTAT	600
35	GGTGATAGTG CCGGTGTCTT GGACGGCCCA CAACATCATC CAAGACTTCT ACAATCCGCT	660
	GGTGGCCTCC GGGCAGAACG GGGAGATGGG TGCCTCGCTC TACGTCGGCT GGGCCGCC	720
40	CGGNCTGCTG CTCCCTGGCG GGGGGCTGCT TTGCTGCAAC TGTCACCCCC GCACAGACAA	780
	GCCTTAATCCTCC GCCAAGTATT CTGCTGCCCG CTCTGCTGCT GCCAGCAACT ACGTGTAAAG	840
	TGCCACGGCT CCACTCTGTT CCTCTCTGCT TTGTTCTTCC CTGGACTGAG CTCAGOGCAG	900
45	GCTGTGACCC CAGGAGGGCC CTGCCACGGG CCACTGGCTG CTGGGGACTG GGGACTGGC	960
	AGAGACTGAG CCAGGCAGGA AGGCAGCAGC CTTCAAGCCTC TCTGGCCAC TCGGACAAC	1020
50	TCCCAAGGCC GCCTCTGCT AGCAAGAACCA GAGTCCACCC TCCCTCTGGAT ATTGGGGAGG	1080
	GACGGAAGTG ACAGGGTGTG GTGGTGGAGT GGGGAGCTGG CTTCTGCTGG CCAGGATGGC	1140
	TTAACCCCTGA CTTTGGGATC TGCCCTGCATC CGTGTGTTGGCC ACTGTCCCCA TTTACATTTT	1200
55	CCCCACTCTG TCTGCCCTGCA TCTCCCTCTGT TGCGGGTAGG CCTTGATATC ACCTCTGGGA	1260
	CTGTGCCCTTG CTCACCGAAA CCCCGGCCCA GGAGTATGGC TGAGGCCCTTG CCCACCCACC	1320
60	TGCCCTGGAA GTGCAGAGTG GATGGACGGG TTTAGAGGGG AGGGCGAAG GTGCTGTAAA	1380

CAGGTTTGGG CAGTGGTGGG GGAGGGGGCC AGAGAGGCCG CTCAGGTGCG CCAGCTCTGT	1440
GGCCTCAGGA CTCTCTGCCT CACCCGCTTC AGCCCAGGGC CCCTGGAGAC TGATCCCCTC	1500
5 TGAGTCCTCT GCCCCTTCCA AGGACACTAA TGAGCCTGGG AGGGTGGCAG GGAGGAGGG	1560
ACAGCTTCAC CCTTGGAAAGT CCTGGGGTTT TTCCCTCTCC TTCTTGTTG TTTCTGTTT	1620
10 GTAATTAAAG AAGAGCTATT CATCACTGTA ATTATTATTA TTTCTACAA TAAATGGAC	1680
CTGTGCACAG GRAAAAAAAA AAAAG	1705

15

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1167 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:
 TGCAGGAATT CGGCAGAGGT TTTCCGCTAG ACTCTGGCAG TTGGTGAGCA TCATGGCAC

CGTTACAGCC ACAACCAAAG TCCCGGAGAT CCGTGATGTA ACAAGGATTG AGCGAACATCGG	120
30 TGCCCACCTCC CACATCCGGG GACTGGGGCT GGACGATGCC TTGGAGCCTC GGCAAGGCTTC	180
GCAAGGCATG GTGGGTCAAG TGGCGGCACG GCGGGCGGCT GGCGTGGTGC TGGAGATGAT	240
35 CCGGGAAGGG AAGATTGCCG GTCGGGCAGT CCTTATTGCT GGCCAGCCGG GCACGGGGAA	300
GACGGCCATC GCCATGGCA TGGCGCAGGC CCTGGGCCT GACACGCCAT TCACAGCCAT	360
CCCCGGCAGT GAAATCTTCTT CCCGGAGAT GACCAAGACC GAGGCGCTGA CGCAGGCCCT	420
40 CGGGGGTCC ATCGGCGTTC GCATCAAGGA GGAGACGGAG ATCATCGAAG GGGAGGTGGT	480
GGAGATCCAG ATTGATCGAC CAGCAACAGG GACGGGCCTC AAGGTGGCA AACTGACCT	540
CAAGACCACA GAGATGGAGA CCATCTACGA CCTGGGCACC AAGATGATTG AKTCCCTGAC	600
45 CAAGGACAAG GTCCAGGCCG GGGACGTGAT CACCATCGAC AAGGCGACGG GCAAGATCTC	660
CAAGCTGGGC CGCTCCCTCA CACGGGCCCG CGAACTACGA CGCTATGGGC TCCCAGACCA	720
50 AGTTCTGCA GTGCCAGAT GGGGAGCTCC AGAAACGCAA GGAGGTGGTG CACACCGTGT	780
CCCTGCACGA GATCGACGTC ATCAACTCTC GCACCCAGGG CTTCCTGGCG CTCTCTCAG	840
55 GTGACACAGG GGAGATCAAG TCAGAAAGTCC GTGAGCAGAT CAATGCCAAG GTGGCTGAGT	900
GGCGCGAGGA GGGCAAGGGG GAGATCATCC CTGGAGTGCT GTCATCGAC GAGGTCCACA	960
TGCTGGACAT CGAGAGCTTC TCCCTCCCTCA ACCGGGCCT GGAGAGTGAC ATGGCGCTG	1020
60 TCCACCAAGGT CTATGGGGAT GCCGTGAGGG CTCTGGTAGC TGGTGCCCCG GATTGCGTGT	1080

--

ATGCCACGGT TGGTGGCCTC GTGCCAATT CCTGCAGCCC GGGGATCCA CTAGTTCTAG 1140
 AGCGCCGCC ACCGGGTGG ANCTCCN 1167

5

10 (2) INFORMATION FOR SEQ ID NO: 108:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1907 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

GGCACAGGGG AATCATCGTG TGATGTGTG GCTGCCTTG TGAGTGTGTG GAGTCCTGCT 60
20 CAGGTGTTAG GTACAGTGTG TTTGATCGTG GTGGCTTGAG GGAAACCCCTT GTTCAGAGCT 120
GTGACTGCGG CTGCACTCAG AGAACGCTGCC CTTGGCTGCT CGTAGCGCCG GGCCCTCTCT 180
25 CCTCGTCATC ATCCAGAGCA GCCAAGTGTCC GGGAGGCAGA AGGTACCGGG GCAGCTACTG 240
GAGGACTGTG CGGGCTTGCC TGGGCTGCC CCTCCGCGT GGGGCCCTGT TGCTGCTGTC 300
30 CATCTATTTC TACTACTCCC TCCCAAATGC GGTOGGCCCG CCCTCACTT GGATGCTTGC 360
CCTCCTGGGC CTCTCGCAGG CACTGAACAT CCTCCTGGC CTCAAGGGCC TGGCCCCAGC 420
TGAGATCTCT GCAGTGTGTG AAAAAGGGAA TTTCAACGTG GCCCATGGGC TGGCATGGTC 480
35 ATATTACATC GGATATCTGC GGCTGATCCT GCCAGAGCTC CAGGCCCGGA TTGGAACTTA 540
CAATCAGCAT TACAACAACC TGCTACGGGG TCCAGTGAGC CAGCGGCTGT ATATTCTCCT 600
40 CCCATTGGAC TGTGGGTGTC CTGATAACCT GAGTATGGCT GACCCCAACA TTGCGCTTCT 660
GGATAAACTG CCCCAGCAGA CGGGTGACCG TGCTGGCATC AAGGATCGGG TTTACAGCAA 720
CAGCATCTAT GAGCTCTGG AGAACGGCA GCGGGGGGC ACCTGTGTCC TGGAGTACGC 780
45 CACCCCTTG CAGACTTGT TTGCCATGTC ACAATACAGT CAAGCTGGCT TTAGCGGGGA 840
GGATAGGCTT GAGCAGGCCA AACTCTCTG CGGGACACTT GAGGACATCC TGGCAGATGC 900
50 CCCTGAGTCT CAGAACAACT GCGCCTCAT TGCCTACCAAG GAACCTGGCAG ATGACAGCAG 960
CTTCTCGCTG TCCCAGGAGG TTCTCCGGCA CCTGCGGCAG GAGGAAAAGG AAGAGGTTAC 1020
TGTGGGCAGC TTGAAGACCT CAGCGGTGCC CAGTACCTCC ACGATGTCCC AAGAGCCTGA 1080
55 GCTCCTCATC AGTGGAAATGG AAAAGCCCT CCCTCTCCGC ACGGATTCT CTTGAGACCC 1140
AGGGTCACCA GGCCAGAGCC TCCAGTGGTC TCCAAGCCTC TGGACTGGGG GCTCTCTTCA 1200
60 GTGGCTGAAT GTCCAGCAGA GCTATTTCTT TCCACAGGGG GCCTTGCAGG GAAGGGTCCA 1260

360

	GGACTTGACA TCTTAAGATG CGTCTTGTCC CCTTGGGCCA GTCATTCCC CTCTCTGAGC	1320
	CTCGGTGTCT TCAACCTGTG AAATGGGATC ATAATCACTG CCTTACCTCC CTCACGGTG	1380
5	TITGTGAGGAC TGAGTGTGTG GAAGTTTTC ATAAACTTTG GATGCTAGTG TACTTAGGGG	1440
	GTTGTGCCAGG TGTCTTCAT GGGGCCTTCC AGACCCACTC CCCACCCCTC TCCCCCTCCT	1500
10	TTGCCCGGGG ACGCCGAAC TCTCTCAATGG TATCAACAGG CTCTTCGCC CTCTGGCTCC	1560
	TGGTCATGTT CCATTATGG GGAGCCCCAG CAGAAGAATG GAGAGGAGGA GGAGGCTGAG	1620
	TTTGGGTAT TGAATCCCCC GGCTCCCACC CTGCAGCATC AAGGTTGCTA TGGACTCTCC	1680
15	TGCCGGCAA CTCTTGCCTA ATCATGACTA TCTCTAGGAT TCTGGCACCA CTTCCCTTCCC	1740
	TGGCCCTTA AGCCTAGCTG TGTATCGCA CCCCCACCCC ACTAGAGTAC TCCCTCTCAC	1800
20	TTGCGGTTTC CTTATACTCC ACCCCCTTCT CAACGGTCCT TTTTAAAGC ACATCTCAGA	1860
	TTAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAGGG CGGCCGC	1907

25

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 611 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

35	ATGAATTAAC GCCAAGCTNT NAATAGGGAC TCACTATGGG GGAAAGNTGG GTAACGCCTG	60
	CAGGTACCGT TCCGGAAATTC CCGGGTCGAC CCACGGGTCC GATGGGGCTT TAGTAAATCA	120
40	GGCTTGCAGG CTCAAAGCTG CAATCTGCC ACTCTCAGGT ACTGAGACTT TGTGGGCTC	180
	AGACACCAGG AAGAAAGTIG GGATACAGTC ATTTGAGTTA AAAAGGGAAT GACCCCTCAG	240
45	AAACCCGCAT TAGCAGTGTG ACTCTTGGAA GTGCCTTAC TTTAACGCT CTCTGTTCTG	300
	AAAAAGAGGT GTTGGTTAC GTGTGAGCCA ACATCACTT TTGTTAGCTG TGATTTACCT	360
	TTGTCGGTTT AAAAGACTTC ACGGAGCCAT TCTGTATACA AGGTGTGCTC TTTCCAATGT	420
50	AGAAGGGTT ATGGAAAAGG GTGGGATCCT TTGCTGTAAA CTGGAGAGAC CAGTCCAAA	480
	CAGAGGGAA TTTTAAGCCC TTCTCATCAC CCAATTGGAT GTTTTGCTT ATAGCAAATT	540
55	CCTGCAAAAT AAATAAATAA ATATTGCAA AACTAAAAA AAAAAAAAAA AAAAAAAAAA	600
	GGGGGGNCCN C	611

60

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 2632 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

10	TCC CAG CTCT CAG GACA AGG GCC CTC GGG CG ATC TTT TAAA AAAC CG GATT GGG TGT CTT T	60
	CT AAA AAT AC AACC AGT ACT TC ATCG TCAA GTT TCT GGG A AGGG AGT CCC CT CCAG ATTC	120
15	TC ATGG AGTG ACA AAAT CTT G ACT CTT GCT C CT GGAA ATT TT TC AGG CCAA ACT AGC GTTT	180
	CT AC AAAT GAT TT ATTT TGG CA AAT TT GTCT T GATT ATGG GT GGCT GAT GAG GAAC GTG CTT	240
20	TT GTT ACC AA CCG AA ACT GG GCG CGG GTGA GGG CGT GTAC GCA ATG AGTC CGGA AGA GAGG	300
	TG AAAT GCT T TCG G TAGG CA CT CCAC CGG CT GTGA AGA TG GG CGG CT GC GT GGCT TCAG	360
	GT GTT GCCT G TC ATT CTCT C GCT TCT GGG A GCT CAC CC CGT C ACC ACT GTC GTT TT CAGT	420
25	GCG GG AC CGG CA ACC GTAG C TGCT GCG AC CGG TCCA AA AT GGC ACAT TCC GATA CC CGT CG	480
	GGG AAAA ATT AT TT TA GTTT TGG AA AGAT C CT CTT CAG AA AT ACC ACT AT CTT CCT GAA G	540
30	TTT GAT GGAG AAC CTT GTGA CCT GTCT TT G AAT ATA ACCT GGT ATCT GAA AAG CGT GAT G	600
	TG TT AC AA ATG AA AT CTATA A CTT CAAGG CA GA AGA AGT AG AGT TGT ATTT GG AAAA ACT TT	660
	AAG GAAA AAAA GAG GCTT GTC TGG AAA ATAT CAA ACAT CAT CAA AAT GTT CC AGA ACT GC	720
35	AG TGA ACT CT TT AAAA ACACA GAC CTT TCT GGAG ATTT TA TGCA TGACT GC CT CT TT TA	780
	GGAG AAAA AC AGG AGG CTA A GGAG AAT GGA ACA AA CCT TA CCT TAT TGG AGAC AAA ACC	840
40	GCA ATG CAT G AAC CATT GCA AACT TGG CAA GAT GC ACC AT AC ATTT TAT TGT AC AT ATT	900
	GG CATT TCAT CCT CAA AGGA AT CAT CAAA GAA AAT TCAC TGAG TA ATCT TTT ACCAT G	960
	ACT GTT GAAG TGA AGGG TCC CT ATGA ATAC CTC AC ACT TG AAG ACT ATCC CTT GAT GATT	1020
45	T TTT TCAT GG TG AT GT GT AT TG TATA ATG TC CT GTT GG GT TG CT GT GG CT GG AT	1080
	GC CT GCT ACT GGAG AGA ATCT CCT GAGA ATT CAG TTT MGGA TT GG GT CT GT CAT CT CC TG	1140
50	CGA ATG CT TG AGAA AGC TGT CT TCT ATG CG GAA ATT CAGA AT ATCC GATA CAA AGG ARAA	1200
	TCT GT CCAGG GT GCT T TGT CCT GAGAR CT GCT T TCT CAG CAG TGA AA AC G CTC ACT GG CT	1260
	CGA ACC CT GG TC ATC ATAGT CAG TCT GGG A TAT GG C AT CG TCA AGCC AC G CCT GG AGT CA	1320
55	CT CT TCATA A GG TGT AGTA GC AGRAG CCC TCT ATC TTT G TCT CT GG C AT GG AAG GGG	1380
	TC CT CAG AGT TACT GGG CC CAG ACT GAT C TT GCT T CTT GG CTT TAT C CC CT GG CT T	1440
60	TC CT AGAC AC TGC CT TGT GC TGG TGG AT AT TATT AGC CT GACT CAA ACA AT GAAG CT AT	1500

	TAAAACCTCG GACGAACATT GTAAAACCTCT CTTTGTATCG GCATTTCACC AACACGCTTA	1560
	TTTGGCAGT GGCAAGCATCC ATTGTGTTTA TCATCTGGAC AACCATGAAG TTCAGAATAG	1620
5	TGACATGTCA GTCGGACTGG CGGGAGCTGT GGGTAGACGA TGCCATCTGG CGCTTGCTGT	1680
	TCTCCATGAT CCTCTTTGTC ATCATGGTC TCTGGCGACC ATCTGAAAC ACCAGAGGT	1740
10	TTGCCCTTTTC ACCATTGTCT GAGGAAGAGG AGGAGGATGA ACAAAAGGAG CCTATGCTGA	1800
	AAGAAAGCTT TGAAGGAATG AAAATGAGAA GTACCAAACA AGAACCCAAT GGAAATAGTA	1860
	AAGTTAACAA ACCACAGGAA GATGATTGAG AGTGGGTAGA AGAGAATGTT CCTTCTCTG	1920
15	TGACAGATGT AGCACTTCCA GCCCTTCTGG ATTCAAGATGA GGAACGAATG ATCACACACT	1980
	TTGAAAGGTC CAAAATGGAG TAAGGAATGG GAAGATTGAG AGTTAAAGAT GGCTACCATC	2040
20	AGGGAAGAGA TCAGCATCTG TGTCAGTCCTT CTGTACGGCT CCATGGGATT AAAGGAAGCA	2100
	ATGACATCCT GATCTGTCCT TTGATCTTTC GGCAATTGGAG TTGGCGAGAG GTGTAGAAC	2160
	AAAGAGAACCA TCTTACTGAA AACAAAGTCA TAAGATGAGA AAATCTACG AGCTTCTTAT	2220
25	TTACAACACT GCTGCCCCCT TICCTCCCAG ACTCTGACAT GGATGTTCAT GCAACTTAAG	2280
	TGTGTTGTC CTGAACTTTC TGTAATGTTT CATTTCCTAA ATCTGACAAA CTAAAAGTT	2340
30	TAACGTCTTC TAAAAGATTG TCATCAACAC CATAATATGT AATCTCCAGG AGCAACTGCC	2400
	TGTAATTTTT ATTTTATTAG GGAGTTACAT AGGTGATGGG GGAAATTGTT AACTACCTTT	2460
	CATTTCCTG GGAAGTCAAG GTTACATCTT GCAGAGGTTG TTTTGAGAAA AAAGGGCCCT	2520
35	TCTGAGTTAA GGAGCCATAG TTCTATCAAT GATCAAAAGA AAAAAAAA AACTCGATCG	2580
	GCACCGAGGGG GGGCCCCGTA CCCAATTGCGC CCTATGGGAN TCGAATGAGA CC	2632

40

(2) INFORMATION FOR SEQ ID NO: 111:

	(i) SEQUENCE CHARACTERISTICS:	
45	(A) LENGTH: 2249 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:	
	GAATTGGCA CGAGCTCACC GTGCTGCGTG ACACAAGGCC AGCCTGCGCC TACGAGGCCA	60
	TGGACTTTKT RATGGCCCTC ATCTACGACA TGGTACTGSW TGTGGTCACC CTGGGGCTGG	120
55	CCCTCTTCAC TCTGTGGGG AAGTTCAAGA GGTGGAAGCT GAAOGGGGCC TTCCCTCTCA	180
	TCACAGCCTT CCTCTCTGTG CTCACTGGG TGGCCTGGAT GACCATGTAC CTCTTCGGCA	240
60	ATGTCAAGCT GCAGCAGGGG GATGCCTGGA ACGACCCAC CTTGGCCATC ACGCTGGGG	300

	CCAGCGCTGG GTCTTCGTCA TCTTCCACGC CATCCCTGAG ATCCACTGCA CCCTTCTGCC	360
5	AGCCCCTGCAG GAGAACACGC CCAACTACTT CGACACGTGG CAGCCCCAGGA TGCGGGAGAC	420
	GCCCTTCTGAG GAGGACGTGC AGCTGCGCG GGCCTATATG GAGAACAAAGG CCTTCTCCAT	480
	GGATGAACAC AATGCCAGCTC TCCGAACAGC AGGATTTCCC AACGGCAGCT TGGGAAAAG	540
10	ACCCAGTGGC AGCTTGGGA AAAGACCCAG CGCTCCGTT AGAAGCAACG TGTATCAGCC	600
	AACTGAGATG GCCGTCGTGC TCAACGGTGG GACCATCCCA ACTGCTCCGC CAAGTCACAC	660
15	AGGAAGAMAC CTTTGGTGAA AGACTTTAAG TTCCAGAGAA TCAGAATTTC TCTTACCGAT	720
	TTGCCTCCCT GGCTGTGTCT TCTTGAGGG AGAAATCGGT AACAGTTGCC GAACCAGGCC	780
	GCTCACAGC CAGGAATTT GGAAATCCTA GCCAAGGGGA TTTCTGTAA ATGTGAACAC	840
20	TGACGAACTG AAAAGCTAAC ACCGACTGCC CGCCCCCTCCC CTGCCACACA CACAGACACG	900
	TAATACCAGA CCAACCTCAA TCCCCGAAA CTAAAGCAAA GCTAATTGCA AATAGTATTA	960
25	GGCTCACTGG AAAATGTGGC TGGGAAGACT GTTTCATCCT CTGGGGTAG AACAGAACCA	1020
	AATTACAGC TGGTGGGCCA GACTGGTGGT CGTTGGAGGT GGGGGGCTCC CACTCTTATC	1080
	ACCTCTCCCC AGCAAGTGCT GGACCCAGG TAGCCTCTTG GAGATGACCG TTGCGTTGAG	1140
30	GACAAATGGG GACTTTGCCA CCGGCTTTGC CTGGTGGTT GCACATTTCA GGGGGGTCAG	1200
	GAGAGTTAAG GAGGTGTTGG GTGGGATTCC AACGTGAGGC CCAACTGAAT CGTGGGGTGA	1260
35	GCTTTATAGC CAGTAGAGGT GGAGGGACCC TGGCATGTGC CAAAGAAGAG GCCCTCTGGG	1320
	TGATGAAGTG ACCATCACAT TTGGAAAGTG ATCAACCACT GTTCCCTCTA TGGGGCTCTT	1380
	GCTCTAGTGT CTATGGTGAG AACACAGGCC CCGCCCCCTTC CCTTGAGAG CCATAGAAAT	1440
40	ATTCTGGCTT GGGGCAGCAG TCCCTCTTC CCTTGATCAT CTGGCCCTGT TCCTACACTT	1500
	ACGGGTGTAT CTCCAATCC TCTCCCATT TTATTCCCTT ATTCAATTCA AGAGCTCCAA	1560
45	TGGGGTCTCC AGCTGAAANS CCCTCCGGGA GGCAGGTGG AAGGCAGGCA CCACGGCAGG	1620
	TTTTCCGGGA TGATGTACCC TAGCAGGGCT TCACGGGTTTC CCACTAGGAT GCAGAGATGA	1680
	CCTCTCGCTG CCTCACAAGC AGTGCACACCT CGGGTCCCTT CGTTGCTAT GGTGAAAATT	1740
50	CCTGGATGGA ATGGATCACA TGAGGGTTTC TTGTTGCTTT TGGAGGGTGT GGGGGATATT	1800
	TTGTTTTGGT TTTTCTGCAG GTTCCATGAA AACAGCCCTT TTCCAAGGCC ATTGTTCTG	1860
55	TCATGGTTTC CATCTGTCTT GAGCAAGTCA TTCCCTTGTT ATTTAGCATT TCGAACATCT	1920
	CGGCCATTCA AAGCCCCAT GTCTCTGCA CTGTTGGCC AGCATAACCT CTAGCATCGA	1980
	TTCAAAGCAG AGTTTTAACCC TGACGGCATG GAATGTATAA ATGAGGGTGG GTCCCTCTGC	2040
60	AGATACTCTA ATCACTACAT TGCCTTTCT ATAAAACCTAC CCATAAGCCT TTAAACCTTTA	2100

AAGAAAAATG AAAAAGTTA GTTTTGGGG CGCGGGGGAG GACTGACCGC TTCATAAGCC	2160
5 AGTACGTCTG AGCTGAGTAT GTTGTATTA ACCTTTGAT ATTTCTAAA AAAAAAAA	2220
AAAANCCCG GGGGGGGGCC CGGCGCTGG	2249

10

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

20 GATACTATAA GGCAAGTGAC TCAAGGGTCG GCCGTTAGAC TAGGGATCC CGGGTGCAGG	60
ATTCGGCAG AGCGCCGCCG GAGCGGAAGT GCTGGCGCCC CGCGGGCCGC TGCGCTCCGG	120
25 GANCCCAAAA TCATGAAATG CACCGTGAG ACCCGGAAGA AAAGGAGGAA TTGGCCGTGC	180
CCGAGAAATAG CTCCGTCAG CTGTTAAGG AAGAAATCTC TAAACGTTT AAATCACATA	240
30 CTGACCAACT TGTGTTGAA TTTGCTGAA AAATTTGAA AGATCAAGAT ACCTTGAGTC	300
AGCATGGAAAT TCATGATGAA CTACTGTC ACCTTGTCAT TAAAACACAA AACAGGCCTC	360
AGGATCATTG AGCTCAGCAA ACATATACAG CTGGAAGCAA TGTTACTACA TCATCAACTC	420
35 CTAATAGTAA CTCTACAGT GGTTCTGCTA CTAGCAACCC TTTGGTTA GGTGGCCTTG	480
GGGGACTTGC AGGTGTGAA AGCTGGGTT TGAATACTAC CAACTTCTCT GAACTACAGA	540
40 GTCAAGATGCA GCGCAACTT TTGCTTAACC CTGAAATGAT GGTCCAGATC ATGGAAAWC	600
CCYTTGTTCA GAGCAGCTC ATGAAATCT GACCTGATGN AGACAGTTAA TTATGGCAA	660
TCCACAAATG CAGCAGTAA TACGAGAAA TCCCAGAAAT TAGTCATATG TTGAATAATC	720
45 CAGATATTAAT GAGACAAACG TTGCAACTTG CCCAGGAATC CAGCAATGAT GCAGGGAGATG	780
ATGAGGAACC AGGACCGAAC TTTGAGCAAC CTAGAAAGCA TCCCAGGGGG ATATAATGCT	840
50 TTAAGGCGCA TGTACACAGA TATTGAGGA CCAATGCTGA GTGCTGCACA AGAGCAGTT	900
GGTGGTAATC CATTGCTTC CTGGTGAGC AATACATCCT CTGGTGAAGG TAGTCACACCT	960
TCCCGTACAG AAAATAGAGA TCCCTCTCCC AATCCATGGG CTCCACAGAC TTCCCCAGAGT	1020
55 TCATCAGCTT CCAGCGGCAC TCCCTGGACT GTGGGTGGCA CTACTGGTAG TACTGCCAGT	1080
GCCACTTCTG GGCAGAGTAC TACTGGCCA AATTTGGTGC CTGGAGTAGG AGCTAGTATG	1140
60 TTCAACACAC CAGGAATGCA GAGTTGTTG CAAACAATAA CTGAAAACCC ACAACTTATG	1200

	CAAAACATGT TGTCTGCCCT CTACATGAGA AGCATGATGC AGTCACTAAG CCAGAACCT	1260
	GACCTTGCTG CACAGATGAT GCTGAATAAT CCCCTATTG CTGGAAATCC TCAGCTTCAA	1320
5	GAACAAATGA GACAACAGCT CCCAACTTTC CTCCAACAAA TGCAGAATCC TGATACACTA	1380
	TCACCAATGT CAAACCCTAG AGCAATGCAG GCCTTGTTAC AGATTCAAGA GGGTTTACAG	1440
10	ACATTAGCAA CGGAAGCCCC GGGCCTCATC CCAGGGTTA CTCCCTGGCTT GGGGGCATTA	1500
	GGAAGCACTG GAGGCTCTTC GGGAACTAAT GGATCTAACG CCACACCTAG TGAAAACACA	1560
	AGTCCCACAG CAGGAACCAC TGAAACCTGGA CATCAGCAGT TTATTCAAGA GATGCTGCAG	1620
15	GCTCTTGCTG GAGTAAATCC TCAGCTACAG AATCCAGAAG TCAGATTICA GCAACAACTG	1680
	GAACAACTCA GTGCAATGGG ATTTTGAAAC CGTGAAGCAA ACTTGCAAGC TCTAATAGCA	1740
20	ACAGGAGGTG ATATCAATGC AGCTATTGAA AGGTTACTGG GCTCCCAGCC ATCATAGCAG	1800
	CATTTCTGTA TCTKGAAAAA ATGTAATTAA TTTTGATAA CGGCTCTTAA ACTTTAAAAT	1860
	ACCTGCTTTA TTTCATTTG ACTCTTGGAA TTCTGTGCTG TTATAAACAA ACCCAATATG	1920
25	ATGCATTTA AGGTGGAGTA CAGTAAGATG TGTGGGTTT TCTGTATTIT TCTTTCTGG	1980
	AACAGTGGGA ATTAAGGCTA CTGCATGCAT CACTCTGCA TTTATTGAA TTTTTAAAA	2040
30	ACATCACCTT TTATAGTTGG GTGACCAAGAT TTGTCCTGC ATCTGTCAG TTTATTIGCT	2100
	TTTTAACAT TAGCCTATGG TAGTAATTAA TGTAGAATAA AAGCATTAAA AAGAAGCAAA	2160
	AAAAAAAAAAA AAAAATTCTT GCGCCCGCGA ATTCTTCT	2198

35

(2) INFORMATION FOR SEQ ID NO: 113:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1043 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

	CTGAAGTGTATGTGAGG AAGAAGAGGC TCCTACTGTATGACAGCTTGTCTACAGAT	60
	CCTCCAGAA ATCTCTGGC CAGGTGGAAC CCAGGGTCAG AGAGGGATGG GAGAGAGGTT	120
50	TAATTTCCA TGATAAAATAA AAATCTATAA AATAATAAAC AAGAGAAAAG AGATTGGAAA	180
	CAGCCAGGTT GGAGCAGTGA GTGAGTAAGG AAACCTGGCT GCCCTCTCCA GATTCCCCAG	240
55	GCTCTCAGAG AAGATCAGCA GAAAGTCTGC AAGACCTAA GAACCATCAG CCCTCAGCTG	300
	CACCTCCTCC CCTCCAAGGA TGACAAAGGC GCTACTCATC TATTGGTCA GCAGCTTCT	360
60	TGCCCTAAAT CAGGCCAGGC TCATCAGTCG CTGTGACTTG GCCCAGGTGC TGCAGCTGGA	420

	RGACTTGGAT GGGTTGAGG GTTACTCCCT GAGTGACTGG CTGTGCCTGG CTTTTGTGGA	480
	AAGCAAGITC AACATATCAA AGATWAATGA AAATGCAGAT GGAAGCTTIG ACTATGGSCT	540
5	CTTCCAGATC AACAGCCACT ACTGGTGCAA CRATTATAAG AGTTACTCGG AAAACCTTIG	600
	CCACGTAGAC TGTCAAGATC TGCTGAATCC CAACCTTCTT GCAGGCATCC ACTGCGCAA	660
10	AAGGATTGTG TCOGGAGCAC GGGGGATGAA CAACTGGGTT AGAATGGAAG KTTGCACTGT	720
	TCAGGCCGGC CACTCTCTA CTGGCTGACA GGATGCCGCC TGAGATKAAA CARGGTGCCG	780
15	GTGCACCGTG GARTCATTCC AAGACTCCTG TCCTCACTCA RGGAATTCTTC ATTTCCTTCTT	840
	CCTACTGCCT CCACCTCATG TTATTTCTT CCCCTCCCCT TTACAACCAA AACTGACCAG	900
	AGCCCCAGGA ATAAATGGIT TICCTGGCTT CCTCCTTACT CCCATCTGGA CCCAGTCCCC	960
20	TGGTTCCCTGT CTGTTATTTG TAAACTGAGG ACCACAATAA AGAAATCTTT ATATTTATCG	1020
	AAAAAAAAAA AAAAAAAACT CGA	1043

25

(2) INFORMATION FOR SEQ ID NO: 114:

	(i) SEQUENCE CHARACTERISTICS:	
30	(A) LENGTH: 703 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:	
	GAATTGGCA CGAGTGGCGG GCCACCACGG CGGTTTTCTG ACGCTGGCGG TGGACGCAGG	60
	CAGCATGGAC CACGGTTGCT GGGGGATGG GGACCGTCTA TGGTCAGTTG CCTTAGAAGT	120
40	GGTGAGATGG GAAGCTGCAG TTGGAAGACC CTGGAGGATG CCTGACAAGG GGATGTCTGA	180
	CACATGATTG GAGCTCTTT TGAAATGTTT CTTGCCCTTC CTGGAGCAGA GGAGCCATTA	240
45	TTTATGCAGG TACATCGAAG TCTTTGACC TCCATACAGT GATTATGCTT GTCATCGCTG	300
	GIGGTATCCT GGCGGCCTTG CTCTGCTGA TAGTTGTGCT GCTCTGTCTT TACTTCAAA	360
	TACACAACGC GCTAAAAGCT GCAAAGGAAC CTGAAGCTGT GGCTGTAAAA AATCACAACC	420
50	CAGACAAGGT GTGGTGGGCC AAGAACAGCC AGGCCAAAC CATTGCCACG GAGTCTTGTC	480
	CTGCCCTGCA GTGCTGTGAA GGATATAGAA TGTGTCAG TTTTGATTCC CTGOCACCTT	540
55	GCTGTTGCGA CATAAAATGAG GGCCTCTGAG TTAGGAAAGG TGGGCACAAA AATCTTCATG	600
	AGCAATACTT CTTAGTAGAT TGTTTGTGTTA TTCAAATCAA GTTCTAGTGT TTTTATGTGA	660
60	GATTATATAA TTTACAGTGT TGTTTTATAT ACTTTGAAAT AAA	703

(2) INFORMATION FOR SEQ ID NO: 115:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3684 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

15	GGCAGAGGGG GCATGAGGAG GAGGAGGATT ACCGCTACGA GGTGCTCACG GCCGAGCAGA	60
	TTCTACAACA CATGGTGGNA ATGTATCCGG GAGGTCAACG AGGTCACTCA GAATCCAGCA	120
	ACTATCACAA GAATACTCCT TAGCCACTTC AATTGGGATA AAGAGAACGCT AATGGAAAGC	180
20	TACTTTGATG GAAACCTGGA GAAGCTCTTT GCTGAGTGTC ATGTAATTAA TCCAAGTAAA	240
	AAGTCTCGAA CACGCCAGAT GAATACAAGG TCATCAGCAC AGGATATGCC TTGTCAGATC	300
25	TGCTACTTGA ACTACCCTAA CTCGTATTC ACTGGCCTTG AATGTGGACA TAAGTTTGT	360
	ATGCAGTGCT GGAGTGATAA TTAACTACC AAAATAATGG AAGAAGGCAT GGGTCAGACT	420
	ATTCGTGTC CTGCTCATGG TTGTGATATC TTAGTGGATG ACAACACAGT TATGCGCCTG	480
30	ATCACAGATT CAAAAGTTAA ATTAAAGTAT CAGCATTAA TAACAAATAG CTTTGTAGAG	540
	TGCAATCGAC TGTTAAAGTG GTGCTCTGCC CCAGATTGCC ACCATGTGT TAAAGTCCAA	600
35	TATCCTGATG CTAAACCTGT TCGCTGAAA TGTGGGCGCC AATTTTGCTT TAACTGTGGA	660
	GAAAATTGGC ATGATCCTGT TAAATGTAAG TGTTAAAGA AATGGATTAA AAAGTGTGAT	720
	GATGACAGTG AAACCTCCAA TTGGATTGCA GCCAACACAA AGGAATGTCC CAAATGCCAT	780
40	GTCACAATTG AGAAGGATGG TGGTTGTAAT CACATGGTCT GTGCTAACCA GAATTGTAAA	840
	GCAGAGTTTT GCTGGGTGTG TCCTGGCCCA TGGGAACACAC ATGGATCTGC CTGGTACAAAC	900
45	TGTAACCGCT ATAATGAGGA TGATGCAAAG GCAGCAAGAG ATGCACAGGA GCGATCTAGG	960
	GCAGCCCTGC AGACGTACCT GTTCTACTGT AATCGCTATA TGAACCACAT GCAGAGCCTG	1020
	CGCTTTGAGC ACAAACTATA TGCTCAGGTG AAACAGAAAA TGGAGGAGAT GCAGCAGCAC	1080
50	AACATGTCTT GGATTGAGGT GCAGTTCTTG AAGAAGGCAG TTGATGTCTT CTGGCAGTGT	1140
	CGTGCCACAC TCATGTACAC TTATGTCTTC GCTTTCTACC TCAAAAAGAA TAACCAGTCC	1200
55	ATTATCTTG AGAATAACCA AGCAGATCTA GAGAATGCCA CAGAGGTGCT CTCGGCTAC	1260
	CTTGAACGAG ATATTTCCCA AGATTCTCTG CAGGATATAA AGCAGAAAGT ACAAGACAAG	1320
	TACAGATACT GTGAGAGTCG ACGAAGGGTT TTGTTACAGC ATGTGCATGA AGGCTATGAA	1380
60	AAAGATCTGT GGGAGTACAT TGAGGACTGAA GAATGCCCT GCATAAAATG AACTCTGAAA	1440

	ACTTTACCAT CTAGAGTGCT CATGCAATTA AAACAAAACA AACACAAACA AGGAGGCAC	1500
5	AAGCCTATTG TGACACCACT GGTCGTAGT ACCAGAATTG TTTTGTAAAT GGAAAGTTA	1560
	AGTAAATTAT ATTGTAATAA AAAGGTAGAT AAACCATTGT ACAACAGTAT TCTAGGCCG	1620
	CAACAAAAGT GTGACAGACA CACTAAAAGC CCTCCAACCT TAACTTGTAA CGTAGCTCA	1680
10	TTCTCAAAGC TGACTCCITT TTTTTCTTTT CCCTTTCTCT GAGTGTAGTA CAGTTAAAAT	1740
	TTCAAAACACC TCCTTGACAC TGCTTTTCAT GTTCAAACCA GCCATTTGT TGTACTTTGG	1800
15	TAAAGGACCT CTTCCCCCTTC CTCCCCTACA CATAACAGATA CACCCACACA CAGACTGACT	1860
	CTCTTTCTCT CATAACCCAA GGTCATGAGT GAATGATGCT TAGTTCCCTG TAAAGAAAAT	1920
	CTTGGGATGG CGAAAGGGT AGGCAGGAAG AGGATTCAAC AAACGAAAAA CATAAAAAC	1980
20	TGTATATGA CTTTTAAAAC AAGAGGACAA CACAGTATTT TTCAAAATTG TATATAGCGC	2040
	ATATGCATGG ACAAAAGCAAG CGTGGCACGT GTTGCATAA TGTTAATTA CAAAAAAATA	2100
25	TTTATTCTTT AAAAATCTTC AAGATTATGT CTATTTGCTG TGCATTTCT TTCAGTTGC	2160
	TTATCTTCC CGGGTTGGGG TTGGGATAAA GGTTGTCGG TTTAGCACCT CTGGAAGACC	2220
	TATCTAGAGC TCTTTCACTT TCCTGAGGTT ATTTTGCCY TTCTGGGTTT GGTATGTCTG	2280
30	TTGCCGGCCA TGGGCTNCAY GCCTTGAATT CCTGCTCTG ATCAGGGACA AGGGAGGTCA	2340
	AGCTCTGACT AATGCCATGA CCTGATTAAG GGGTACAGCA GGGAGTTTG TTGCTACAGC	2400
35	TCATGAATTA ACCTGTCCCA ACCTAATCCC CCTCCATGGC ATCATGCCCTC TACCCAAGCC	2460
	TTTGTGTGCC CATGTTATGC ACACAGCTGT AGGCATTCTT AAGTCCCCTG TCGCATCCAG	2520
	TGGAAGCATT TAAAAATTTC TTTTACTTTT TGGTTTCCCT TAAATTGCTG CTTTTCAGAT	2580
40	TTTAGTTATG GCTCGTCTGC TCACCCCTTC TCTACATTAG GGTGTCAAAG AGAATGTTT	2640
	GCTTTAAATA TAAATAGCCA TTCATTTAGT CTCAGATTGT GAATTAAAA TGGTGGATAC	2700
45	CGAAATTGCT TGTGTGGTT GCTGTGGTT TGGTTGAAG GCAAACACCC CTAGAACATG	2760
	ATATTCCCAT CTAGTGCATT TAAATAGAAA TCACTGAGTT TGCTGCTTT TTATTGTCAG	2820
	CAGATAGGAG AATTAATAAT GCATTTAGC TGTGATGTCC ATTTTATGA AATTCTACT	2880
50	AAGAGCTATG TTAAAAGTAA AGGATGGTGG TGGTTGTATT AACTATATAC CTGTTTAGGC	2940
	CATTCTGGCT GTGGTATTTT TCAATAGGTC AGCATCTGTA AATCTGTCAAG TTTTATACAG	3000
55	GAGTGCAGAG TGAACTAGGC AACTAGATTA AGAGGTCTAA ATATGAAATA CCAGTTGAGG	3060
	CTGAGGACCT CCTCGTCTTC CTTAAATGT CTTTGCCTA GGGAGTGTG ACCATTGTG	3120
	AGGCAGCTTT GTCTGCTCTT ACACTGTACA TCCTATTACT CCATTGGAA GTAGGTTCAC	3180
60	TTCCCTCTGG CCTTTTGCTT AAGTTAGGCT TTGCTGAATC AACCTACTT TTCCCTTTAG	3240

	AAAAGGTTGT TACAGGAGAT TTACTGGCAA CTGTTCTTTT CCCATCAAAA ATCAGTGAAT	3300
5	GTTGCTGAG TATAAAATGCT GCTTCCTTAA ACCACTTGTC GCTTTAGGAT CAACTTTACC	3360
	TGTACCTTTT CTCCCTTCCT CCCCTGCCAC CTCAGGTGCA AATCTGAAC T CAGTGTCTGC	3420
	TTCTTCCATT TTCTCGTCTC TCTCCCCCTCT TCCCCCATTA TCCATATGAC ATTATTTTAC	3480
10	TTCAAATGAC AGCATCAATC TAAAAAAGAT ATACATTAAA ACTAAGGAGT TTTTTAAAG	3540
	AAAGCCTGAA TAAGTTCCCTT TCCCTGGTAA CTTTGAAAAG CAGTCAGAGT TGCTATATAG	3600
	ATATATGTGG CTCCCTTAAA ATGCTTGTG TATGTGTGGT GTTTAAAAAA AAAAAAAAAA	3660
15	TTGGGGGGGG GGCCCGGTNC CCAT	3684

20

(2) INFORMATION FOR SEQ ID NO: 116:

	(i) SEQUENCE CHARACTERISTICS:	
25	(A) LENGTH: 1965 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:	
30	AAGAAAGGGT ATTAAAATTC TAGATCACAT ATGGACCCGG GAAGGTTTTT NACCCCTCTGT	60
	TAGTGACATC GAGTCTCCCA CTAGACAAAA TAGGTGGAAA AATCTCTCGA GGGCTCACAT	120
35	TGTTTTGTCA TCTTCAGGAA AAACACCACC AGGCCATACC ACAGCCTGCC CAGTGAGGCG	180
	GTCTTGCCA ACAGCACCCG GATGCTGGTG GTGGCCTTTG GGCTGCTGGT GCTCTACATC	240
40	CPTCTGGCTT CATCTTGGAA GCGCCCAGAG CCGGGGATCC TGACCGACAG ACAGCCCCCTG	300
	CTGCATGATG GGGAGTGAAG CAGCAGGAAG GGGCTCCCAA GAGCTCTGG TGGTGCAGCC	360
	TGIGCTCCCC TCAGAACGTC TGCTCTTCCC AGGGCTCCCG GCTGGTTCA GCAGGGCACT	420
45	TTCTTCCAAT GCTGGGCCCA GACTTCCTGC CTGGGTGCTG GCCTGCCCTC TCCGGNCCGC	480
	TTGCTGCTTG TCTGCTTTCC TTGGTGGYTT TGCTGGGTGC TGGGCTGCC CTCTCCGGCC	540
	GCTTGCTGCC TGTCTGCTTT CCTTGGTGGC TTTGCTGGGT GCTGGGCCTG CCTTCTCTGG	600
50	CTGCTTGCTG CCTGTCCTGCT TTCCCTGGTG GCTTTGGCTT CTGCACTCT TGCGTCASC	660
	TCTCAGGTCC TCCATTACACA CGAGGTCTTC CTCGCTCTGG CCGCTCTTGC TGCTCTGTG	720
55	TGAAGAWATC AGACTGATTT CCTCTTAAGA CTCCCTAGGGA TGTTGGTGAAG AGCTGGGACT	780
	CAAGTGCAGT CCACGGTGTG AACATGAGG GARGTGAGGT GTCCGTCCAC TTCCCCCATA	840
60	AAGGTGTGCA TTTCAGTTAG GCTGCCCGC CACAGAGGAG GCTTCATCTG CTCTGCCATC	900

	CAGCCCCATC TGGATGTGAG GTGGGGTGGA GACATCATGG GGTGATTGCA GAAAGGGGA	960
	GTGGCGGCCCG ACAGCAGCTTC TGCTGAGGAG CTGACCGCTC TGAGCTGTTG TGTTCTGTAT	1020
5	TGCTGCTCTG TGTCTGCATG TATTGTGACC GTGGGGCTCC ACCTCTTCCA GCTGCTGCTA	1080
	CAGCTGAGGC CTGGATCCCG GCCTTTCCCT GTGACTTAAG TGCTGTAC CGGCANGCAG	1140
10	CCCTACAAAT CCTGGTGACC TGCTCTCCCA AGAACAGAGC CTGTCCTCAG ATGTCCTCAGT	1200
	ACCGATGAGT AACAGAGGTG GCTGTGGACT TCCTCTACTT CTCCCTGCTG GATCAGGGCC	1260
	TTCCCTGCCTC CCGCTGGCA GGCTCTGGCCT TGCTCTCTTG GCAGGGCCCC AGCCCCTCTG	1320
15	ACCACTCTGC AGCTCACCAT GCAGCTGATG CCAAAGTTGT GGTGTCCAGT GTGCAGCAGC	1380
	CCTGGGAGCC ACTGCCACCT TCAGAGGGGT TCCTTGCTGA GACCCACATT GCTTCACCTG	1440
20	GCCCCACCAT GGCTGCTPTGC CTGGCCCAAC CTAGCGTTCT GTGCCATGCT AGAGCTTGAG	1500
	CTGTTGCTCT TCITCAGGGG AGGAAATAGG GTGGAGAGCG GGAAGGGCT TGCTCTAAAG	1560
	TGTTGCTGCT GTGGCTTTTG TGCCCTCTCC AAAGACGCAC TGCCAGGTCC CAAGCTTCAG	1620
25	ACTGCTGTGC TTAGTAAGCA AGTGAGAAGC CTGGGGTTTG GAGCCCCACCT ACTCTCTGGC	1680
	AGCATCAGCA TCCTACTCCT GGCAACATCA GGCCAACGTC CACCCAGCC TCACATTGCC	1740
	AGATGTTGGC AGAAGGGCTA ATATTGACCG TCTTGACTGG CTGGAGCCTT CAAAGCCACT	1800
30	GGGATGTCCT CCAGGCACCT GGGTCCCCTG ACCAGCTCCC CGTCTCCATA GGGGTAGGCA	1860
	TTTCACTGGT TTATGAAGCT CGAGTTTCAT TAAATATGTT AAGAATCAAA GCTGTCTTGT	1920
35	TTCAGGCTGC TATAACAAAA ATATAATAGC CTGGGTGGCT TAAAC	1965

40 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 503 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

50	AGTGATCCCC TTGCTCGGC CTCCAAAAT GCTGGAATTG TAACCGTGGG CCTCTGCACC	60
	CGGCCTGGTC CGCAATTAA AAACGCACAG CCACCATTC CTYTCAGAA AGCACCCAGA	120
	TGCCCTTGGG AGAACCAAGCC TCCTCCATGG AGGAAAGCTT GGGATCTGCC TTCCCACCTG	180
55	GGGAGGAGAG GGATCTGTGG AAAATCCTTC TGACGGACTT CCCCTCAGTG CCTGATCCAT	240
	ACTCAATAGT AGAAAAGTA AGAAATATAC AAAGATAGCA GATACACGGA GACAGTTCCC	300
60	CAAATAGCTG ACCGAWTAGC GCAGAAGCAA TATTGAAGAC CTAATAGCTG AGACATTTC	360

AGAACTGATA AAGTGCATCC AGCCACAGAT CAAGCAGCCC AGAAAATTCC AGGCAGCATIC 5 AACAAATAAA TAGCCCCACA TGCACCCGTG AAAATGCAGA AGACCAAACA AAAAAGTCGG GTCAACAGCC AGAGTTAAAG AGG	420 480 503
------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------

10

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS: 15 (A) LENGTH: 1133 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:	
20 GGCACAGCTT GGAATGAACC CCTGTGGATA AGGGGGACTA TTAGATAGAA TAAACATCAA TAAATGCTTG ATGAATAAAC GCTAATCCTA CCTTCCCAGC CTGACACCTC CCAGTGGACA 25 CCACACITCA CTTGAACGCT TAGAAACCTT TCCCACCCAT GCTTCCAGCC CTGGCTTCAT GTTGCCATTT CTCACCCCCA GAACAGGCCG CCCGCCTGAA GAAACTACAA GAGCAAGAGA 30 AACAAACAGAA ACTGGAGTTT CGTAAAAGGA TGGAGAAGGA GGTGTAGAT TTCATTCAAG ACAGTGGGCA GATCAAGAAA AAGTTTCAGC CAATGAACAA GATCGAGAGG AGCATACTAC ATGATGTGGT GGAAGTGGCT GGCGTGACAT CCTTCTCCCTT TGGGAAGAT GATGACTGTC 35 GCTATGTCAT GATCTTCAAA AAGGAGTTTG CACCCTCAGA TGAAGAGCTA GACTCTTACC GTCGTGGAGA GGAATGGGAC CCCAGAGG CTGAGGAGAA CGCGAACNTG AAGGAGCTGG 40 CCCAGAGGCA ANGAGGAGGA GGCAAGCCAG CAGGGGCCTG TGGTGGTGAG CCCTGCCAGC GACTACAAGG ACAAGTACAG CCACCTCATC GGCAAGGGAG CAGCCAAAGA CGCAGCCCAC ATGCTACAGG CCAATAAGAC CTACGGCTGT KTGCCCGTGG CCAATAAGAG GGACACACGC 45 TCCATTGAAG AGGCTATGAA TGAGATCAGA GCCAAGAAGC GTCTGCGGCA GAGTGGGGAA GAGPTGCGC CAACCTCTTA GGCGCCCCGC CCAGCTCCCT TTGACCCCTG GGGCAGGGCA 50 GGGGGCAGGG AGAGACAAGG CTGCTGCTAT TAGAGCCAT CCTGGAGCCC CACCTCTGAA CCACCTCTTA CCAGCTGTCC CTCAGGCTGG GGGAAAACAG GTGTTTGATT TGTCAACCGTT GGAGCTTGGGA TATGTGGCTG GCATGTGTGT GTGTGTGTGA GAGTGTGAAT GCACAGGTGG 55 GTATTAAATC TGTATTATTC CCCGTTCTTG GAATTTCTT CCCATGGGGC TGGGGTACTT TACATTCAAT AAATACTGTT TAACCCAAAA AAAAAAAA AAAAGAAAGA AGN	60 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900 960 1020 1080 1133

(2) INFORMATION FOR SEQ ID NO: 119:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

	GGGCACAGCT GAAGCTGCAG ACCTCCCCAG GGGATGGCTC CTCTCCCCA GGAGCCCCGA	60
15	GCGAGGGGAG GCAGAAAGCC TGGGCTCTGG GGGGTGGCCT GCGGACAGCT GTGCTGTGGG	120
	CCGGGGGCTG GGCCTGTCCC ACAGGGNCGT GGAGCTCGTG GTTCTGAGCA GCCAGCTGGG	180
	TGGTGTCTGG GGATAAGCTGG GAGGCACAGC GGCTGCCATG TGGGACTGGG ACTGGAGTGC	240
20	TCCCTGGTCT TGGCCTCTGT GGCTCAGCCT TGCTCTGGTC TGCCTGAGTG CAGGGCCAA	300
	GGGGCACAGG GCCAGTGAGG CCGGCCACGC TCGGGCCCTC ACCTGTGAGA TGGGGTCGGA	360
25	ATTTKACACA GCCTANGGCT TGTTCTTGG TKGTNGAMCG TGGACTYCTK AGAACGGAG	420
	TGCTGGTCCT .GAAAGGGGTG GTGGGAGACC AGCTGCTTTT CTCGCTGTTT TTCTCTAGG	480
	AGATTAACAA AAAACAGAAA GCACAAGACG AACTCAGTAG CAGACCCAG ACTCTCCCT	540
30	TGCCAGACGT GGTTCCAGAC GGGGAGACGC ACCTCGTCCA GAACGGGATT CAGCTGCTCA	600
	ACGGGCATGC GCCGGGGGCC GTCCCAAACC TCGCAGGGCT CCACCGAGGC AACCGGCACC	660
35	ACGGACTCCT GGGTGGGCC CTGGCGAACT TGTTTGTGAT AGTTGGTTT GCAGCCTTTG	720
	CTTACACGGT CAAGTACGTG CTGAGGAGCA TCGCGCAGGA GTGAGGCCA GGCGCCGAGA	780
	CCCAAGGGCG CACTGAGGGC ACCGCGCACC AGAGCGTGAC CTOGGCAGGC TGGACACACT	840
40	GCCCAGCACA GGCAGACCCA CCAGGCTCCT AGGTTAGCT TTTAAAAACC TGAAAGGGGA	900
	ACCAAAAACC AAAATGTGTG ACTGGGCTTT GGAGGAGACT GGAGCCTCAG CCCTGTCTG	960
45	GCCACGGGCC GCTGGGGCTG GTGTGGGTGG GCCTTGTGTG CTGGATTGT AGCTTATCIT	1020
	CCGTGTGTGTC TTTGGACCTG TTTTAGTAAA CCCGTTTTTC ATTTAAAAA AAAAAAAAAA	1080
	AAACTTTGGG GGGGGGCCCC N	1101

50

(2) INFORMATION FOR SEQ ID NO: 120:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AGCTTCTCTG TCCAGTCTTG AACTCTGGGS TCTCTTGGAA CTTCCTCAC CCCTCTCAGC	60
5 CTGAATATTG CTTCCATGGA TTCCACTCAA CCAGACTTTG GATCTGTGCC TACTTAATCA	120
ACCTTATCTT TGCAATAATGT TCGGGCCAC CTTCCACTCC TTGGTCTTG TTCCTCTTG	180
10 GCCTAACCTG TCCCTCTCC ACTTCACATC CCCGGTGGGA CAGCATTCT CCTTCCTCCC	240
AACCTCCCTC CGTCTCARAA AAAAAAAAAA AAAAAAAAAA TT	282

15

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2635 base pairs
- 20 (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:	
-----------------------------------------------	--

TAAGGGGGTG TGTGCTCACC TCCCTCTGAC CCTTAACACT CCTGCTCTGC CCAGACCAAC	60
AGAGAGAGCT GTCCCTGAGA CCCCGGAGAG AAGCAGCTGC CGAAAGCTGC AGCCTTCCG	120
30 CACTCTGAGA CCATGATCTT CCTCCTGCCA GGGGAGAGCC ACCCACAGGC CATGTCAGC	180
CCCACTTCCC TCAGCCCCA GGYTTCCCTT CTGGCCCTTC TGAGGATTCC CTAGGGCTGC	240
35 CCCGCAGAGG GGYTTCCCCA AGCTCTGTTT TGAAGCCTGC AATGTGGAAA AGTGAGAAAGT	300
CAGAGGGAAC AGGACAGGTG CAGCCGGCT CTGAGGCAC ACCCACACCC TCGCTGTTCC	360
CCAACATCCC CTGAGCAGTG TGAGCTCATC TCACCAGATG AGAAGAGGCC CTGTGCATT	420
40 YTTTTGTTTG TTTGGTGTG TTTCCCCCA CCCATCCAGT TCTCCTCAGC AAAGCAAATT	480
CCTTAACACC TTTGGTGGAG AATTCTTAC CCAGACTTGG GGCTGTGATG CCCTTCAGTG	540
45 CGTGGTGAGT GCAGCGTGTG TCGGTGTGCC TGTGTGTGAA CCTGGGGGCC ATCCTGGTGG	600
CCTGGGAGCG TGAGGAGAGG CCCCTGTGT GCTGGGTGAG TGGTGGGTGT GGGGTCAATG	660
50 CAGTGAGGCT CTCTGGGTGA GGCTCCCAAC CTGGCAGTCC CCAGCCTCCC AGCATCTGTG	720
AGCGTCTGTT GGACTTTACA GAAGAGCCTC ATCCYGTCTG CCCCTCACTC TGCCCTGGAA	780
TCAACATCTT CCGAGTCCTT CTTGGGGAA ATAGCAGAGC CCCACTTAAC TCCATAAACT	840
55 GCTTOCCATT CGCGAGCCA GTTCTGATTG TTGAGGTGTC GCGTCGTTC AGGTCCCCCA	900
GTCCTCTCTT TCTCCTGTCC TCTCTCTGTC CTTCACCTCC CCACCTCCAGC CCCGGCTCAG	960
TTCAGGGAAA TGCTGTCCA YATCAGCCCT CTGCTCTCTG AGGCAGCCGC CCCTCTGACT	1020
60 CGGAGCTACT TGAAAATTCT GCTCTTGCTA GGATTGGAGT CTACCTATCT CTTCCATTG	1080

	TCCCCAGCTGG AGTTCTGGAA CTTTCCTCTT CGGGGTGCGGG GTGGGGGTTC TTACGGATCC	1140
5	TGGGGGGCCG GGGGAAGGAA GGAGTTCAGA GGAAGGGTGT CCCCTGGCTT CTTCCTGTC	1200
	CCCTCCGCTC CTGGGACACG TGCTCTCTCT GTCTCTGGGT CTTCTGGGT TGCACGGTTC	1260
	TGTGTCCCTG TAAATATGTT TTAGGAAGAA AGCAAAAAGG ACTGAATAG CCTCTGGTAG	1320
10	GATTGCAGGG GTCCAGCCTT GCCTGTTTCC GAAGCCCCCA CACTGCCCTT CGCCCCACTG	1380
	AGACTGGTCC CCTCAAAAGG TAGACAAAAAC AGCAGCTCCC TGTGGAGTTG AAGGGCGGCC	1440
15	TCAAAGTGGC TTTTGTTAG ACAAGGTAA GGTTCTCPA TGACGAATG TGCTGATCGG	1500
	TCCTTCCTCA GCTCTTGTAT TTGTGACCTT GACCAAGGGG CCTGCCAAC AGCCCCCTCCA	1560
	GTGCCCTCTC CTGATGCGT CGCTCCCTCC TGCCCCCACT CCCCTGGTT AGGGAGGTAG	1620
20	GGGAATTAGG GCCATGCTGG AAGAACCTTA ACCATGTGTT CAAGAAGG TTTCTTGCTT	1680
	GCTTGGTCCT GGAACCTCCC TTGGCTGCC CAGGCCTCTT TGGCCCHGG GTGGTGGGG	1740
25	AGGTGGATGT CAGATCTGGT AGGTGGCAGC AGAGAAATA AATGTGCTT GAGAGACAC	1800
	TCAGAGAGGG TCCAAGGGTG ATGGAGAAGG AAGCTGCCC TGGGAGCTTG GAGGGARGG	1860
	GTGGTGGGTG CGGGCATCTT GACTGCCCTC TGTGGTCCA CACGTGGGG GTGGTCACCC	1920
30	CYCTTCACTC CAGCCCCCT GCCTTCAGCC TCCATGAGC TTCACCTCTT TCCAACTCA	1980
	CTTGGAGGG CGTGGGGTCC GTGGCATCA ACACGGGAC CCTCTGCTTC ACCAAGGCC	2040
35	GAGCCCTCAG CCCCTGGGA GAACAAATGG CTGAGCTTGT ATACCTGGGG TCGTCGAGAG	2100
	GCTGGGGCT GCGGGCAGTC CCAGGGGAGA GACACCAAG AAGGAGLCC AGCTCTCCG	2160
	AGGAAGTTCC CAGCAGAGCA AACTGCTTTC CAGCCTGAG CCTGCTAAA CTGTTGATG	2220
40	TGCAATAACT GAGCTTAGAG TTAGGAATTG TGTCAGTG CTTGGATTC CGCTGTAGA	2280
	TTAACTGCT GAAATTGTAT CTCTCAGTAA TTTTGTAGT CTTTAAATA ATTGAAAC	2340
45	AAAGTGTAG ACTGTGTGCG TGTGCGTTGA TGGGCACTCA AGAGTCCCT GAGTCATCCA	2400
	GCCTGCCTT TCCCCGCCC CCCCATCTC TCACGTCCCG CCCGGCCCTC ACTGGGGAC	2460
	CCTGCCTCGT GTCGTCCTTA TCTGCCTATT ACTCAGCTA AGGAAACGG TACACTCCAC	2520
50	ACATGCATAA AGGAAATCAA ATGTTATTT TAAGAAAATG GAAATATAA ACTTTATAAA	2580
	CACCAAAAAA AAAAAAAAAA ACCCNNGGGG GGGGCCGTA ACCCATTTGG CCTAA	2635
55		

(2) INFORMATION FOR SEQ ID NO: 122:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 994 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GAATTCCGGCA GAGGTTGGC GAAGATAGGG AATAAGGAAG CACAGGAGTA	60
AAGCACAGGA GTAGGGGAGA TATACAGCGG TCAGGATAAG GGGAAACGG CGGTGGTTGC	120
10 SCAAGAGGTG AAACAAGATG TGAGAGACAA GGGTAGGGA AGAAATGGGG CACCGGTTAG	180
GTTCAGAACG GCATAGACCG TGGCGGACGG GCAATGCGAG CCCACAGAA AGGAACGTGAG	240
15 GGGTGGGCTA TTITAARGGA GATGGTCCTT CAGCCCTCTT YTTTCTGCG TAGTTCTCCT	300
CCTCCAGGCC GCGCGCGGAT ATGTCGTCCG GAAACCAGCC CAGTCTAGGC TGGATGATGA	360
20 CCCACCTCCT TCTACGCTGC TCAAAGACTA CCAGAATGTC CCTGGAATTG AGAAGGTTGA	420
TGATGTCGTG AAAAGACTCT TGTCTTTGGA AATGGCCAAC AAGAAGGAGA TGCTAAAAAT	480
CAAGCAAGAA CAGTTATGA AGAAGATTGT TGCAAACCCA GAGGACACCA GATCCCTGGA	540
25 GGCTCGAATT ATTGCCTTGT CTGTCAGAT CGCAGTTAT GAAGAACACT TGGAGAAACA	600
TCGAAAGGAC AAAGCCCACA AACGCTATCT GCTAATGAGC ATTGACCAGA GGAAAAGAT	660
30 GCTCAAAAC CTCGTAACA CCAACTATGA TGTCTTGAG AAGATATGCT GGGGGCTGGG	720
AATTGAGTAC ACCTTCCCCC CTCTGTATTA CCGAAGAGCC CACCCCGAT TCGTGACCAA	780
GAAGGCTCTG TGCAATTGGG TTTTCCAGGA GACTCAAAAG CTGAAGAAGC GAAGAAGAGC	840
35 CTTAAAGGCT GCAGCAGCAG CCCAAAAACA AGCAAAGCGG AGGAACCCAG ACAGCCCTGC	900
CAAAGCCATA CCAAAGACAC TCAAAGACAG CCAATAAATT CTGTTCAATC ATTTAAAAAA	960
40 AAAAAAAAAA AAAAAAAAAA AAAAAGGGGA GGGG	994

45 (2) INFORMATION FOR SEQ ID NO: 123:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1542 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear .

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

55 GGCASAGCCA CCTCGGGCCC GGGCTCCGAA GCGGCTCGGG GGCGCCCTTT CGGTCAACAT	60
CGTAGTCCAC CCCCTCCCCA TCCCCAGCCC CGGGGGATTC AGGCTCGCCA CGGCCAGCC	120
AGGGAGCCGG CGGGGAAGCG CGATGGGGGC CCCAGCCGCC TCGCTCCTGC TCCCTGCTCCT	180
60 GCTGTTCGCC TGCTGCTGGG CGCCCGGGGG GGCCAAACCTC TCCCAGGACG ACAGCCAGCC	240

	CTGGACATCT GATGAAACAG TGGTGGCTGG TGGCACCGTG GTGCTCAAGT GCCAAGTGAA	300
5	AGATCAGGAG GACTCATCCC TGCAATGGTC TTAACCCITGC TCAGCAGACT CTCTACTTTG	360
	GGGAGAAGAG AGCCCTTCGA GATAATOGAA TTCAGCTGGT TAMCTCTACG CCCCCACGAGC	420
	TCAGCATTCA CATCAGCAAT GTGGCCCTGG CAGACGAGGG CGAGTACACC TGCTCAATCT	480
10	TCACTATGCC TGTGCGAACT GCCAAGTCCC TCGTCACTGT GCTAGGAATT CCACAGAACG	540
	CCATCATCAC TGGTTATAAA TCTTCATTAC GGGAAAAAGA CACAGCCACC CTAAACTGTC	600
15	AGTCCTCTGG GAGCAAGCCT GCAGCCCGGC TCACCTGGAG AAAGGGTGAC CAAGAACTCC	660
	ACCGAGAACCC AACCCGCATA CAGGAAGATC CCAATGGTAA AACCTTCACT GTCAGCAGCT	720
	CGGTGACATT CCAGGTACCC CGGGAGGATG ATGGGGCGAG CATCGTGTGC TCTGTGAACC	780
20	ATGAATCTCT AAAGGGAGCT GACAGATCCA CCTCTCAACG CATTGAAGTT TTATACACAC	840
	CAACTGCGAT GATTAGGCCA GACCCCTCCCC ATCCTCGTGA GGGCCAGAAG CTGTTGCTAC	900
25	ACTGTGAGGG TCGGGCAAT CCAGTCCCCC AGCAGTACCT ATGGGAGAAG GAGGGCAGTG	960
	TGCCACCCCT GAAGATGACC CAGGAGAGTG CCCTGATCTT CCCTTCCTC AACAAGAGTG	1020
	ACAGTGGCAC CTACGGCTGC ACAGCCACCA GCAACATGGG CAGCTACAAG GCCTACTACA	1080
30	CCCTCAATGT TAATGACCCC AGTCCGGTGC CCTCCTCCTC CAGCACCTAC CACGCCATCA	1140
	TCGGTGGGAT CGTGGCTTTC ATTGTCCTCC TGCTGCTCAT CATGCTCATC TTCCCTGGCC	1200
35	ACTACTTGAT CGGGCACAAA GGAACCTACC TGACACATGA GGCAAAAGGC TCCGACGATG	1260
	CTCCAGACGC GGACACGGCC ATCATCAATG CAGAAGGCGG GCAGTCAGGA GGGGACGACA	1320
	AGAAGGAATA TTTCATCTAG AGGGCCCTGC CCACTTCCTG CGCCCCCCCAG GGCCCTGTGG	1380
40	GGACTTGCTG GGGCGTCAC CAACCCGGAC TTGTACAGAG CAACCCGAGG GGGCGSCCCT	1440
	CCCGNITGTT CCCAGCCCA CCCACCCCT TGTTACAGAA TGTYTKGTTT GGGGTGCGGT	1500
45	TTTGTWATTG GTTNGATN GGGGAAGGGA GGGANGCCGG GG	1542

(2) INFORMATION FOR SEQ ID NO: 124:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1390 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

60

CAAGCTCTAA TACGACTCAC TATAGGGAAA GCTGGTACGC CTGCAGGTAC CGGTCCGGAA 60

	TTCCCGGGTC GACCCACGCG TCCGGGCCTC AGGGTGGACG CATGGTCTG CACTGAGGCC	120
	CTCGTCATGG TGGCGCCTGT GTGGTACTTG GTAGCCGCGG CTCTGCTAGT CGGCTTTATC	180
5	CTCTTCCTGA CTCGCAGCCG GGGCGGGCG GCATCACCGGCC AGCAAGAGCC ACTGCACAAT	240
	GAGGAGCTGG CAGGAGCAGG CCCGGTGGCC CAGCCTGGC CCCTGGAGCC TGAGGAGCCG	300
10	AGAGCTGGAG GCAGGCCTCG GCCCGGGAGG GACCTGGCA GCCGCCTACA GGCCCAGCGT	360
	CGAGCCCAGC GGGTGGCCTG GCCAGAACCA GATGAGAACG AGGAGGAAGC TGTCACTCTA	420
	GCCCAGGAGG AGGAAGGTGT CGAGAAGCCA GCGGAAAYTC ACCTGTGGG GAAAATTGGA	480
15	GCTAAGAAC TCGGAAANN GGAGGAGAAA CAAGCGCAA AGCCCAGCK TGAGGCAGAG	540
	GAGGCTGAAC GTGARGWCGG GAAACGACTC GAGTCCCAGC GCGAATGAGT GGAGAGAAGGA	600
20	GGAGGACCGG CTTCGCCTGG AGGAGGAGCA GAAGGAGGAG GAGGAGAGGA AGGCCCCGGA	660
	GGACCAGGCC CAGCGGGAGC ATGAGGAGTA CCTGAAACTG AAGGAGGCCT TTGTGGTGG	720
	GGAGGAAGGC GTAGGAGAGA CCATGACTGA GGAACAGTCC CAGAGCTTCC TGACAGAGTT	780
25	CATCAACTAC ATCAAGCAGT CCAAGGTGT GCTCTTGAA GACCTGGCTT CCCAGGTGGG	840
	CCTACGCACT CAGGACACCA TAAATCGCAT CCAGGACCTG CTGGCTGAGG GGACTATAAC	900
30	AGGTGTGATT GACGACCGGG GCAAGTTCAT CTACATAACC CCAGAGGAAC TGGCCGCCGT	960
	GGCCAACCTTC ATCCGACAGC GGGGCCGGGT GTCCATGCC GAGCTTGCCC AAGCCAGCAA	1020
	CTCCCTCATC GCCTGGGGCC GGGAGTCCCC TGCCCAAGCC CCAGCCTGAC CCCAGTCCTT	1080
35	CCCTCTTIGGA CTCAGAGTTG GTGTGGCTA CCTGGCTATA CATCTTCATC CCTCCCCACC	1140
	ATCCTGGGGA AGTGATGGTG TGGCAGGCA GTTATAGATT AAAGGCCTGT GAGTACTGCT	1200
40	GAGCTTGGTG TGGCTTGGTG TGGCAGAAGG CCTGGCTAG GATCTAGAT AAGCAGGTGA	1260
	AATTTAGGCT TCAGAATATA TCCGAGAGGT GGGGAGGGTC CCTTGGAAAGC TGGTGAAGTC	1320
	CTGTTCTTAT TATGAATCCA TTCATTCAAG AAAATAGCCT GTTGCAAAAAA AAAAAAAAAA	1380
45	AAAAACTCGA	1390

50 (2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1288 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

60	GGCGCGCGGG TGAAAGGCC ATTGATGCCAG CCTGGGGCGG CCTCGGAGCG CGGGGGASCA	60
----	-------------------------------------------------------------------	----

	GACGGCTGACC ACGTTCCCTCT CCTCGGGTCTC CTCCGCCTCC AGCTCCGCCG TGCCCCGGCAG	120
5	COGGGAGCCA TCGGACCCCA GGGCCCCGCC GCCTCCCCGC AGCGGCTCGG CGGCCTCCTG	180
	CTGCTCCCTGC TGCTGCAGCT GCCCGCGCCG TCGAGCGCCT CTGAGATCCC CAAGGGGAAG	240
	CAAAGGGCGC ATCCGGCAGA GGGAGGTGGT GGACCTGTAT AATGGAATGT GCTTACAAGG	300
10	GCCAGCAGGA GTGCCTGGTC GAGACGGAG CCCTGGGGCC AATGGCATTC CGGGTACACC	360
	TGGGATCCCA GGTCGGGATG GATTCAAAGG AGAAAAGGGG GAATGCTGA GGGAAAGCTT	420
15	TGAGGAGTCC TGGACACCCA ACTACAAGCA GTGTTCATGG AGTTCAATTGA ATTATGGCAT	480
	AGATCPTGGG AAAATTGCGG AGTGTACATT TACAAAGATG CGTTCAAATA GTGCTCTAAG	540
	AGTTTGTTAGTAC TTGGCTAAA ATGCAGAAAT GCATGCTGTC AGCGTTGGTA	600
20	TTTCACATTTC AATGGAGCTG AATGTTCAAGG ACCTCTTCCC ATTGAAGCTA TAATTTATT	660
	GGACCAAGGA AGCCCTGAAA TGAATTCAAC AATTAATATT CATGCACCTT CTTCTGTGGA	720
	AGGACTTTGT GAAGGAATTG GTGCTGGATT AGTGGATGTT GCTATCTGGG TTGGCATTG	780
25	TTCAGATTAC CCAAAAGGAG ATGCTCTAC TGGATGGAAT TCAGTTCTC GCATCATTAT	840
	TGAAGAACTA CCAAAATAAA TGCTTTAATT TTCAATTGCT ACCTCTTTTT TTATTATGCC	900
30	TTGGAATGGT TCACTTAAAT GACATTTAA ATAAGTTAT GTATACATCT GAATGAAAAG	960
	CAAAGCTAAA TATGTTACA GACAAAGTG TGATTICACA TGTGTTAAAT TCTAGCATT	1020
35	TTCAATTGTC TTCAATCAAAGTGGTTTCA ATATTTTTTT TAGTTGGTTA GAATACTTTC	1080
	TTCATAGTCA CATTCTCTCA ACCTATAATT TGGGAATATT GTTGTGGTCT TTTGTTTTT	1140
	CTCTTAGTAT AGCATTTTTA AAAAAATATA AAAGCTACCA ATCTTGTAC AATTGTAAA	1200
40	TGTTAAGAAT TTTTTTATA TCTGTTAAAT AAAAATTATT TCCMACAACCC TTAAAAAAA	1260
	AAAAAAAAAA AAAAAAAAAA AAAANAAA	1288

45

(2) INFORMATION FOR SEQ ID NO: 126:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1517 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

AGTGGCTTAA AGGCATCGTT TTAGGGATTA CTGGGAAGTA TCTTCAAAGT AATACATGAG	60
AAACATTCCCT TCCTAAATCC TTATTATATAT TGAATATCGT ATTAAATTGGT TTTCAGAGGT	120

60

	TAAATTAACC ATGTATTCC GCAATAAAATG TCACTTGINT CTTGTATATA ATCTTTTTTA	180
	TATATTACCG GATTGATTCA TTAGTATTTT GTTGAGGATT TTGAGTCTA TATTCTAAAG	240
5	AGATGCTGGT CTGCAGTTT CTTTTTTGT GATAATCTGG TTTTGTATC AGTAATACAG	300
	GCCCCATGAA ACGAGTGGG AAGTGTAC CTCCTTGTA TTTTTCAAG AGTTTGTGAA	360
10	GAATTGCTAT TAATTCTTTA AATGTTGGT AGAACATACC ATTGAAATCA TGTGCTCTGG	420
	GCTTTTTT GAGGGAAAGTG TTCTGATAAC TAATTCAAGTA TCTACTTTT ATAGCTCTGT	480
	TCAGATTTG CTTCTTCTG AGTTAGTTT GGTAATTGT GTATCTCTAG GATTTGTCC	540
15	ATTCATTTA TCTCATTGT TGGCATAAAT TAAACTAAAT TTGGCCTGAG CCTACCTGTA	600
	TATCTTGAGT CCCTCTGAA CGAACTGTAG CCTAACTTGT ACATAAACAA ACTGAAATCC	660
20	TAAATTAGGA ATGTAGTTT TGTAACAGCT CCTGAGTCTC AGGCAGTCAC ACCAGYCAAG	720
	TCTGTCAATT GCAGGCTGCT AACTAAGCAG CCCATGSTCA AATGAGGCAA AAACCTTGC	780
	TTTAACACA TAGTATAGCT TTGTAATCCT TTCTTGCAC ACTCGGGTAA TTCTTCCCT	840
25	TTTCATTCCC KGWATTTCC AKGAATATGA RTCTYCCTTT TTCCCTCTCC TGTCACTCTA	900
	GCTAATGGTT TGTCAATTGT GTTGATCTTT TGAARAACAA ACCTTTGGTT CCACTTCTT	960
	GTTGCATATG CTGARTATTTC TCATAATTGG AGTGGAAAGC TGATCTTGA TTACTTATTT	1020
30	TACTTAGGGC TGAGGAGTTC ATGGACTTCG CAAAACCTCC TTGAATCTAA ATTGCATCTT	1080
	CTTCCCTGGT TTCTGGCTG AACATGTTT TTCCCATCT WANAWACCCT TGGCTTTTC	1140
35	ATKGGCGATT AAGACTAGAG AAAGTTCTAG ATMCCTGTC CTTTATGCT GTCATTTGT	1200
	TTAAAGGCTT TCTATGTAGT AAAACTATCT ATATAGACAA AATAGAGCCT TGAGTTGTGG	1260
	TCTTGAATTG GATCAACATG ATTTACCAACA TTCTGTACTG GATATTCTT CACCTGCTGC	1320
40	TACTGTAAAC CATTATTTC TTGGATCTTC TGTAGAGTAT ATTATCACAG GTACTTTTA	1380
	CAGGGGTGTC TAATCTTTG GTTCCCTGG GCACATTGAA AGAAGAAGAA TTGCTTGGG	1440
45	CCACACATCA AATACGCTAA CACTAATAAT AGTTGATGAG CTAAAAAAA AAAAAAAAG	1500
	GCAAAAAGN CCCAAA	1517

50

(2) INFORMATION FOR SEQ ID NO: 127:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

	TGAAATCTATT CTTTGAACAT TCTACAACAA GAATTACATT ATACTGTTAT ACCAGAGTAC	60
5	TTCTGCAGTG TGAAATAGAT TCGTTTGGAA AATGAACCTG GCTTGTCTAT AAATTACATT	120
	CACAGGCCCTT TTTGCAAATG TGTAACCTGC CTATCAAAGT AGTTGTAGG GCAAATGCAG	180
	AAATATATGTC TCCATCTGGT AAAGTACCTT WTAYTCATGT GGGAAATCAA GTAGTATCAG	240
10	AACTTGGTCC AATAGTCCAA TTTGTTAAAG CCAAGGGCCA TTCTCTTAGT GATGGGCTGG	300
	AGGAAGTCCA AAAAGCAGAA ATGAAAGCTT ACATGGAATT AGTCAACAAT ATGCTGTTGA	360
15	CTGCAGAGCT GTATCTTCAG TGGTGTGATG AAGCTACAGT AGGGRMGATC ACTCATGMA	420
	GGTATGGWTC TCCTTACCCCT TGGCCTCTGW WTCATATTTT GGCCTATCAA AACAGTGGG	480
	AAGTCAAACG TAAGNTGAAA GCTATTGGAT GGGGAAAGAA GACTCTGGAC CAGGTCTTAG	540
20	AGGATGTAGA CCAGTGTGT CAAGCTCTCT CTCAAAGACT GGGAACACAA CCGTATTTCT	600
	TCAATAAGCA GCCTACTGAA CTTGACGCAC TGGTATTTGG CCATCTATAC ACCATCTTA	660
	CCACACAAATT GACAAATGAT GAACTTCTG AGAAGGTGAA AAACATATAGC AACCTCCTTG	720
25	CTTCTGTAG GAGAATTGAA CAGCACTATT TTGAAGATCG TGGTAAAGGC AGGCTGTCA	780
	AGAGTTATGT GTTACTCTCA GGAGTCTTAA CTTTTGAAAT ATGTTTTACT TGAATGTTAC	840
30	ATTAGATATT GGTGTCAGAA TTTTAAAACC AAATTACTGC TTTTGAAAC CTCAAATTAT	900
	ATAATGTATC TTAATGTAATGT GCTTTATATT GTTATTGTG TATACATTAA AATAATTCTG	960
35	AATTATTTAA TCTGATATGT TGTATTCTGT ATCTTGAAAT TTTTGTTCG TTGAAACATG	1020
	CATGCATTTA AAAATAAAGC TTAAACAACT GTAAAAAAA AAAAAAAA CTC	1073

40

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

50	CAACCCCTGC CTTTTTTTG TTTTCCATTG GCTTGGTAGA TCTTCCTCCA TCCCTTTATT	60
	TTGAGCCTAT GTGTGTCTCT GCCCGTGAGA TGAGTCTCCT GAATACAGCA CACTTACTGG	120
55	TCTTGACTCT GTATCCAATT TGCCAGTCTG TGTCTTTCAT TTGGAGCATT TAGCCCATT	180
	ACATTTAAGG TKAATATTGT TATGTGTGAA TTTRATCYTR TCATTATGWT GTTAGCTGGT	240
60	TATTTTGCTT GTTAGTTGAT GCAGTTCTT CCNGGCATCA ATGGCTTTA CAANTTGGCA	300

(2) INFORMATION FOR SEQ ID NO: 129:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1275 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

15	GGCAGAGCCT GTCCCTGCTG CCCCTGCAAA AAAAACCCCC TCTGGTGTGA GCAGGATGGT	60
	TGGAGGTTAT GTGAGCTCCT TCTCCTTTCC TCCAGTTTCC TCTTCCCTTC TCCTCCCTGC	120
	CTCTTTGCT TTTCCTTTTC TTCTGGTAC CCCCTGCCCA TTCCCTGTATT TTCTCCCATC	180
20	GCCATTCTCC CCTCTOCCAC TGTCCTAAC CCGTTCAAAC TCTTCCCTCT TAAATGGTGT	240
	AGATTTCTC TCACCAAGCA CACCCCAGTA TTAATTAAAC TAGCTGCAAA CAGGCAGCAA	300
25	GTTGGTCTACC ATGACAGATG GGTTTTGTGT GTGTGTGTGT GTGTGTAAATT GTAATAAAC	360
	ATATTGARTC ACTCAATAAA CACAGAGTGT CTACTACATG TATCARGCAC TATCATAGAT	420
	GCTAATTAAC GAAACTGAAA TGGCCAGGCC CTCACAGTGG CTCATGCCTA TAATCCCAGC	480
30	ACTTTGGGAG GATGAGGCAG GAGGATCACT TGAGGCCGGG AGTTCAAGAC CAGCCTGGC	540
	AACATAGTAA GACTCCATCT CTACAAAAAA AAAATTTTTT TTATTATACT TTAAGTTTG	600
35	GGTTACATGT GCAGAACGTG TAGTTTGTGT ACATAGGTAT ATACGTGCC CGGTAGTTG	660
	CTGCACCCAT CAACCCATCA CCTACATTAG GTATTTCTCC TAATGTTACC CCTCTCCAG	720
	CCCCCCACCC CGTGACAGGC CCTGGTGTGT GATGTTCCCC TCCCTGTGTC CATGTGTCT	780
40	CATTGGTCAA CTCTCACCTA TGGAGTGAGA ACATGTGGTA TTTGGTTTTC TGATCTTGTG	840
	ATAGCTTGTGCT GAGAATGTKG GTTCCAGCT TTATCCACGT CCCTGCAAAG GGCATAAACT	900
45	CATCCCTTTT TATGGCTGCA TAGTGTCCA TGGTGTATAC GTGCCACATT TTCTTAATCT	960
	ATCATTGATG GACAAGTTTT GCTATTGTGA ATAGTGCCAC AATAAACATA CGTGTGCGTG	1020
	TGTCTTTATA GCAGCATGAT TTATAATCCT TTGGGTATAT ACCCAGTAAT GGGATCACTG	1080
50	AGTCAAATGG TATTTCTCGT TCTAGATCCG TAAGGAATTG CCACACTGTC TTCCACAATG	1140
	TTTGAACTAA TNTACACTCC CACCAACAGT GTAAAAGTGT TTCTATTTC CCACAACCTC	1200
55	TCCAACATCT GTTATTTCT GACTTTTAA TGAACGTCA TCTAACTGGC GTGAGATGGT	1260
	ATCTCATTGT GGTTT	1275

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 472 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

10	CNGAAACCCC GTGAACCCTC CCCGGGTAA AAAGCCCCC CTAAATGGG GGAACGCYTC	60
	ACACGTTATA AAAAACCACT AGAATGTTT GAAAGCGAGA AACAAACAGCT GTGTAGGGTA	120
15	GCTAGCAGTT AGTGTGTAC AGAAGACAGA TATTTGTGCA TTTTGCATT TTCTAAGTTT	180
	GCTGCAATGA GCATGTATTA CTTTCATAGT TATAAAACAC ATGCAAATG CCCTTTAAA	240
20	ATGAAAAAAA ATCCATGAGT GTAAGTGATA TATATGCTTT GGAAAGCCTG GGACGGTCAT	300
	TGTTTACTCT CAATAGTATG TGTTTGCCTT TGCTTTTTG AGACATTTTG TTTTAATCTG	360
	TTGATGACAA TAACCTGTTG ATAATATAAC TTGATAACAA ATAAAATGAC TTATGATTGA	420
25	AWMAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA NN	472

30 (2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 1950 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

40	ACCTCTCAGA ATCTTCTCTC AGCAACCTGA GTCTTCGCCG TTCTCAGAG CGCCTCAGTG	60
	ACACCCCTGG ATCCCTCCAG TCACCTCCCC TGGAAATTCT GCTGTCCAGC TGCTCCCTGT	120
45	CCCGTGCCCTG TNATTCCCTG GTGTATGATG AGGAAATCAT GGCTGGCTGG GCACCTGATG	180
	ACTCTAACCT CAACACAACC TGCCCCCTCT CGCCCTGCC CTTTNTGCC CTCCTCAGTG	240
	TCCAGACCNT TGATTCCCGG CCCAGTGTCC CCAGCCCCAA ATCTGCTGGT GCCAGTGGCA	300
50	GCAAAGATGC TCCGTCCCT GGTGGTCCTG GCCCTGTGCT CAGTGACCGA AGCTCTGCCT	360
	TGCTCTGGAT GAGCCCCAGC TCTGCAACGG GCACATGGGG GGAGCCTCCC CGCGGGTTGA	420
55	GAGTGGGGCA TGGGCATACC TGAGCCCCCT GGTGCTGCGT AAGGAGCTGG AGTCGCTGGT	480
	AGAGAACGAG CCCAGTGAGG TGCTGGCGTT GCCTGAACIG CCCCTCTGCC ACCCCATCAT	540
	CTTCTGGAAC CTTTGTGGT ATTTCACAG GCTACGNCTG CCCAGTATTG TACCAAGGCC	600
60	GGTGCTGCC TCCGTGATG GCCCTTCGMA CTCCCAGGCC CCATCTCCTT GGCTAACCCC	660

	TGATCCAGCC TCTGTTCAAG TACGGCTGCT GTGGGATGTA CTGACCCCTG ACCCCAATAG	720
5	CTGCCCACCT CTCTATGTGC TCTGGAGGGT CCACAGCCAG ATCCCCCAGC GGGTGGTATG	780
	GCCAGGCCCT GTACCTGCAT CCCTTAGTTT GGCACTGTG GAGTCAGTGC TGCGCCATGT	840
	TGGACTCAAT GAAGTGCACA AGGCTGTGGG GCTCCTGCTG GAAACTCTAG GGCCCCCACC	900
10	CACTGGCTG CACCTGCAGA GGGGAATCTA CCGTGAGATA TTATTCCTGA CAATGGCTGC	960
	TCTGGGCAAG GACCACGTGG ACATAGTGGC CTTCGATAAG AAGTACAAGT CTGCCTTTAA	1020
15	CAAGCTGGCC AGCAGCATGG GCAAGGAGGA GCTGAGGCAC CGGGGGGGCGC AGATGCCAC	1080
	TCCCAAGGCC ATTGACTGCC GAAAATGTTT TGGAGCACCT CCAGAACGCT AGAGACCTTA	1140
	AGCTTCCCTC TCCAGCCTAG GGTGGGGAAAG TGAGGAAGAA GGGATTCTAG AGTTAAACTG	1200
20	CTTCCCTGTT GCCTTCATGG AGTTGGGAAAC AGGCTGGAA GGATGCCAG TCAAAGGCTC	1260
	CAAGCGAGGA CAACAGGAAG AGGGATCCAC TGTTACCAAA AGTCCTGATT CCCCCATCAC	1320
	CAACCTACCC AGTTTGTTCG TGCTGATGTT GGGGGAGATC TGGGGGGAGT TGGTACAGCT	1380
25	CTGTTCTTCC CTTGTCCTAT ACCGGGAACCT CCCCTCCAGG GTACCCACAG ATCTGCATTG	1440
	CCCTGGTCAT TTTAGAAAGTT TTTGTTTTAA AAAACAACTG GAAAGATGCA GAGCTACTGA	1500
30	GCCTTGCCTC TGAATGGGAG GTAGGGATGT CATTCTCCAC CAATAATGGT CCCCTTCCC	1560
	TGACGTGCT GAAGGAGCCC AAGGCTCTCC ATGCCCTTCT ACCTAACGTGT TTGTATTTA	1620
	TTTTAAATTAA TTTATCTGG AGCCACAGCC CCCITGCTTA TGAGGTTCTT ATGGAGAGTG	1680
35	AGAAAGGGAA CGGAAATAGG GCACCATGGT CGGGTGGTTT GTAGTCCTT CAAAGTCAGG	1740
	CACTGGGAGC TAGAGGAGTC TCAAGCTCCC CTTAGGAAGA ACTGGTGCCTC CCTCCAGTCC	1800
40	TAATTTTCT TGCCCTGCCGC GCCTTGGGA ATGCCCTCACC CACCCAGGTC CTGACCTGTG	1860
	CAATAAGGAT TGTCCCTGC GAAGTTTGT TGGATGTAAA TATAGAAAA GCTGCTTCTG	1920
45	TCCTTTCAA AANAAAAAAA AAAAAAAACT	1950

(2) INFORMATION FOR SEQ ID NO: 132:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

60

TGGAAGATT AAAATAGGTT TCATATTCT CTTGAATATG AATATATAAG CTTGAATAAG

60

	CTTGAGTCC TATTATATG AATTTTCCT TATTATTCT ACCAATGCTT CTTATATTAA	120
	AGCCTGAATC TTTCGAAATT AGTATATGTA CATTAGCTGC CTGTGGATTA ACATTTCCAT	180
5	GAATGTTT TTTCGATGT TTGATCTTAA ACTTTTGIG TCTTTATATA AGGTATGCTY	240
	CTTTAACCA TGATATTTT ACCACAATA GTGAAAGAC AATCTYCACC TTTTACTTGT	300
10	ATATTTACAT GTPATGAAAT TTTGATGCA TATTACGCT TATTATTTAA CCAACCTATT	360
	TTATTTATC TAGGGCAATT TCGAAAGC CTTATTTCT TGTATTAATC AAATATTTT	420
	AACATGTAAT TTCCYCTAT TAGTAGKAA TACGKTACYC YAAATATATA TTGTGGSTAT	480
15	TTTCAGAATT GCATATGCC TCCTTAATTT ATTAGAGGCT AACCTAAATT ATTACTTTA	540
	CCACTTACTT GAAATTCCTG GAACTTACA ACATTTATTG TTTTATGCAT TTTAATTCTA	600
	CTTGTATTT TACTACTCTT AACATTATT ATTGTTTAG ACAAGCCAA ATATATNTG	660
20	TTATTAACCTT ATTCCTGATT TCTTCTGTA TTTTATGCC ACTATGTATG CTCATTTCC	720
	TTCATGTGA TGAACCTAAT TCGTACTTT TGTTTTTAA TCTGTGCAGG TAGCCTGGCC	780
25	ATTAATTT TTTTGGT TTGCTGAAA AATGCTGTTT ATTTCTATAT CCATACTTAT	840
	GCATATGAA TCTAGTGG ACATTTTTT AGTATTTATA AATGPAAGT CATTWATKG	900
30	GCTCTATCA TTTCGTRGA GAAATCAATT GTCAGCCAA TAGTTTTCA TTTTAAATTA	960
	CNGAATTTT TCGTGTCTG GGTTTAGGA	990

35

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1720 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

45	GTCGATAAG CGACTGTGGT TATCCCCTA AAGTTTACTT CAGCACTAAC ACTAGTGCTT	60
	CGCGTGGAGT TTGCACTTTT CCAGCTTTAT ACAGGATTT CCTTGTACTG GAAGAGTC	120
50	GGATATAGAG ACTCACTAGT G-CATTATTATT GTACAACATC AAGGGAAATA GGATACTCAT	180
	CAACTGGGA TTATTCTTAT CAAACATGG TCTTCPTTGA ATAAGAAAAA TACATAGT	240
	GTTATTATGG ACTTAAACT GYGTAAATG GATATTCTGA TAAAATATT GCTGCTCTG	300
55	AGAGTGTGGG AATATCTGAGA ATATTAGCTT TACTCATCTT GAGCTTGAG GATGTTCTCT	360
	GTACGCCGAT GGTTCTAT TAACTAAAAA AGCTGGTAT TGTAAAATCT CATTTATAAA	420
60	AACTCAGATG AGGAGAAAT TTCTTTGAT GGTGAGACTG TTGTCTTAGT TCAGGAAATT	480

	ATTTAATAAT CCTTTGTTAC CTGTGAATGA AGGAACCTTG TAATTCTGAT TTATCGTAA	540
5	ACATGAGCCT TTCCAGAGTC AGCTTAGACA CTGTTGTCGC AAATAGCCAT GCTTTGCCTT	600
	ATGCCAAGGA GCCCCAGAGG GAGGGCCTAG TCTTCCTCTG TTGCTGTACA TATATTGAAA	660
	TGCTTTTTT TTTTATTTG CATTGTTAT CTATAATGAG CTTTCTGAGC CCTGATATT	720
10	TGTGAGACAA ACAGGAGTTA TTGATGTTAT ACACTCCCTT CCATTCAGGA TTTTCTGCTT	780
	GGAGGGAAAT ATGTTGACCT TAGAGAATTG TGAATATTGT TGCAATTCTT GAATATATTA	840
	CCATGTGAAT AATAGAGACT GTGTTGCTCT CTAGTATAAG CTATATTAT TTTTGATTCA	900
15	TTTGAATTAC TAGTTATAAC TGGAGAAATT TTGTTACCTC TATCCTGGCT TGCCTGACTG	960
	GCTGTATAAT ACCAGCAGCC TCCTTCTAGAG CATCTTAATG AAAACATGGA TGAAAGGAAT	1020
20	TAATGATGAT ATCTGCAGAC TGGTAGAAA ATGGCTTTG TTCCCAGCGT TAACATTTTC	1080
	TTCTCAATCA CATTCAATG TTTGTGGAGA GTGGCAGATT CACACCAGAA ACACTAGGTG	1140
	TTCATATCCA TAGCATGGAT GCAGAATAAG CAGTTGGAG AGAACGTTCT TCCTACCTGG	1200
25	TACTCCTCCC ATTCACTCA GCCCAGCCCC AGACAGGGT TAGCATTCAAG TGTGGGCCCT	1260
	CAGGCAGCCC TGAAGCCTGG CTGGGTCACTC AGATGGGGC AGCCTGTGAC GGGCACCAAGC	1320
30	GGCCTGATTC CAGGGAAAGAG TTCCTGGAGG GTGTTGGCTG TTTTGTGTTAG CTCAGTTTT	1380
	TTCTGGGCTC CACCAATTCTT AACTCCAGGT AGACAAGATA GATGTCACAC ACAACAATT	1440
	TAAAGTATTT TGCTTAGTGC ATTTTGTGTTA TGATTGCACT GTTTGTTCT TATTTAATAG	1500
35	GCTTTTACT TCATTCTATT AAATTTAGT GTTTAGAAGA GGCGGGTACT GTCACTGNGT	1560
	AAAATATGTA ATATTTTATA TGTTATACCA TGTCAATATAT ACTTGAATA TCAGACCTTG	1620
40	CATTCAATAT ACAATGCAAT TGACTCTTGT CAGACCTGCA TTTTCACTG AACAAATAAAA	1680
	AGATTGTCTG GCACCTCAA AAAA AAAAAA AAAAAA AAAAAA	1720

45

(2) INFORMATION FOR SEQ ID NO: 134:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 705 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

	GGCACGAGGC CATCTGGCT CATTCAAGCAG GAAATAATGG AAAAAGCTGC AATATCCAGG	60
	TGTTTACTAC AATCTGGAGG CAAGATCTTT CCTCAGTATG TGCTGATGTT TGGGTTGCTT	120
60		

	GTGGAATCAC AGACACTCCT AGAGGAGAAT GCTGTTCAAG GAACAGAACG TACTCTTGG	180
	TTAAATATAG CACCTTTAT TAACCAGTTT CAGGTACCTA TACGTGTATT TTGGACCTA	240
5	TCCTCATGCC CCTGTATAACC TTTAACCAAG CCAGTGGAAC TCTTAAGACT AGATTTAATG	300
	ACTCCGTATT TGAACACCTC TAACAGAGAA GTAAAGGTAT ACGTTTGTA AATCTGGAA	360
10	GACTTGACTG CTATTCCATT TTGGGTATCA TATGTACCTT GATGAAGANG ATTAGGTTGG	420
	GATACTCAA GTGAAGCCTC CCACTGGAAA CAAGCTGCAG TTGTTTAGA TAATCCCATC	480
	CAGGTTGAAA TGGGAGAGGA ACTTGTACTC AGCATTCAAG ATCACAAAAG CAATGTCAGC	540
15	ATCACAGTAA AGCAATGAAG ACCAGTTTTC CAATGAAAAC TGTGTAATAA GAGCATCAC	600
	AAGTACAAAA TTCTTGTCTT AATTAGTGGG GGTATATAAA AATTCCCTGT AATGGTCAA	660
20	TATTTTTAA AATTGACATT AATAAAGCAT ATTTAAAAG TTTCT	705

25 (2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 323 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

AGCACACACC TCCTTAGTT GCTCCTAAGG TCAATGTCAA CATTGTTGGA GTGCATTTC	60
TGCTCAGGGA GCTTCCCCAG ACCCGGAATG TTTGGTGCTC ACAGACYCTG GCAAGGATCG	120
GTATTGCTGT TCCTCAGTT TGCTGGGGA AATGGAGGST CAGTGACGTT CAGTGACGTG	180
40 CCCAGAGTCA TGCCATTGGC GGGTGGCCCA GKGMTCCAGG TCTCCAGCAC CCCTCGGCC	240
CCTCCCTCACC AGGTACACATC ATCTCCTGGA TTAGAACCTG CTCACATAGT CTGTCCTGAA	300
AGGAAAAAAA AAAAAAAAAA AAC	323

45 (2) INFORMATION FOR SEQ ID NO: 136:
 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 582 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

GGACGGAATG GTGCAACCCCT CCTWAMTTT CTKKGKGTGT TGACAACAGA GGGAGGGAGG	60
-------------------------------------------------------------------	----

	GAAAACATTT TTYGTGGGAG AATCCTACYT CTGCAGSGGA GCCCTTAAGC GATKGATTTC	120
	GAATCTKGAC CCTTTACCAA CTAATTTGAA AGGAAGATAC CTTGGAAATA TTTGGCATTC	180
5	AGTGGGTTAC TGAAACAGCA TTAGTGAATT CATCTAGAGA ACTCTTTCAT TTATTCAAGC	240
	AACAACGTAA CAACCTGGAA ACCTTGTTAC AGTCCAGTTG TGATTTGGG AARGTATCAA	300
10	CTCTACACTG CAAAGCAGAC AATATTAGGC AGCAGTGTGT ACTATTCTC CATTATGTTA	360
	AAGTTTCAT CTTCAAGGTAT CTGAAAGTAC AGAATGCTGA GAGTCATGTT CCTGTCCATC	420
	CTTATGAGGC TTTGGAGGCT CAGCTTCCCT CAGTGTGAT TGATGACCTT CATGGATTAC	480
15	TCTTGTATAT TGGACACCTA TCTGAACCTC CCAGTGTAA TATAGGAGCA TTTGTAAATC	540
	AAAACCAGAT TAAGGTTGAA CTGGTTTCAT TTGATTTTA AG	582

20

(2) INFORMATION FOR SEQ ID NO: 137:

	(i) SEQUENCE CHARACTERISTICS:	
25	(A) LENGTH: 1021 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:	
	TTCGGCAGAG CCCTTGGCGCG CTCTTGAATA CCTGCKTTCT GTAGCGCTAG TTCTCTCAA	60
	GATTTGCTTA GTGTCATTTTC ATTTGGTTT CTTTCTCGC CATGTTTTTC TGTCGGAATT	120
35	ACGGTTCGTT TTGGTTCTAT GTACTCTCTA AAATGTTATC GTTTTCATT TGTCTACTAA	180
	TTTCGTGCA TTGGTTACTA CTGAGTTCT TAATATCTGA CTGGCCTCCG CCCACGGGCT	240
40	CTGCAGANCA TAAAATACTC AGGCTGATGG TAGTGCAGAG ACTCTCCCTC CTTGATCAGC	300
	GCAAACGTTG GTCTGAGGCT TGAGGGATGG AGCAACATTT TCTTGGCTGT GTGAAGGGGG	360
	CTTGGGATTC CGCAGAGGTG GCGCCAGAGC CCCAGCCTCC ACCTATTGTG AGTTCAAGAAG	420
45	ATCGTGGGCC GTGCCCTCTT CCTTTGTATC CAGTACTAGG AGAGTACTCA CTGGACAGCT	480
	GTGATTTGGG ACTGCTTTCC AGCCCTTGCT GGCGGCTGCC CGGAGTCTAC TGGAAAACG	540
50	GAECTCTCTCC TGGAGTCCAG AGCACCTTGG AACCAAGTAC ACCGAAGCCC ACTGAGTCA	600
	GTGGCCGGG GACACAGAAG CAGCAAGARG CACCGTAGA AKARGTGGGG CAGGCAGARG	660
	AACCCGACAG ACTCAGGCTC CRGCAGCTC CCTGGACAGAG TCCTCTCCAT CCYTGGGACA	720
55	GACAGCAGGA CACCGAGGTC TGTGACAGCG GGTGCCCTTT GGAAACGCCGC CATCCTCCTG	780
	CCCTCCAGGC GTGGCGCCAC CTCCCCGGTT TCTCAGACTG CCTGGAGTGG ATTCTTOCGG	840
60	TTGGTTTGC CGCGTCTCT GTACTCTGGG CGTGCTGTC ACGGATCTGT GGAGCTAAGC	900

AGCCCTAGAT AGCAGCAGAA GGCTTTTGG ATTCTCCCTCC TTGAAAAGAT TCTCAGTTAC	960
CAAACGTCTC CACCTAGAAA ATAAAAATAC ATTAAGATGT TGAAAAGAAA AAANAAAAAA	1020
A	1021

10

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1777 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

CGGAAGATGA TGGCTCAAC AGATCCATTG ATGAAGTGAT ACTAAAAAT ATTACTTGGT	60
ATTCAAGAACG AGTTTTAACT GAAATCTCCT TGGGGAGTCT CCTGATCCTG GTGGTAATAA	120
25 GAACCATTCA ATACAACATG ACTAGGACAC GAGACAAGTA CCTTCACACA AATTGTTTGG	180
CAGCTTTAGC AAATATGTCG GCACAGTTTC GTTCTCTCCA TCAGTATGCT GCCCAGAGGA	240
30 TCATCAGTTT ATTTCTTTG CTGCTAAAA AACACAACAA AGTTCTGGAA CAAGCCACAC	300
AGTCCTTGAG AGGTTCGCTG AGTTCTAATG ATGTTCCCTCT ACCAGATTAT GCACAAGACC	360
TAAATGTCAT TGAAGAAGTG ATTCAATGA TGTTAGAGAT CATCAACTCC TGCCTGACAA	420
35 ATTCCCTTCA CCACAACCCA AACTTGGTAT ACGCCCTGCT TTACAAACGC GATCTCTTGT	480
AACAATTTCG AACTCATCCT TCATTTCAAGG ATATAATGCA AAATATTGAT CTGGTGATCT	540
40 CCTTCCTTAG CTCAAGGTTG CTGCAAGCTG GGAGCTGAGC TGTCAGTGGG ACGGGTCTG	600
GAAATCATTA AGCAAGGCGT CGTTGCGCTG CCCAAAGACA GACTGAAGAA ATTTCCAGAA	660
TTGAAATTCA AATATGTGGG AGAGGAGCAG CCCGAGGAGT TTTTATCCC CTAIGTCTGG	720
45 TCTCTTGTCT ACAACTCAGC AGTCGGCCTG TACTGGAATC CACAGGACAT CCAGCTGTT	780
ACCATGGATT CCGACTGAGG GCAGGATGCT CTCCCCACCCG GACCCCTCCA GCCAAGCAGC	840
50 CCTTCAGTT CTTTTATTTG TGGTAACAG AAGTAGACAG ACAGGTTACT TGGTGTATCT	900
TCTGTTAAAG AGGATTGCAC GAGTGTGTTT TCCTCACACA CTTTGATTG GAGAATTGGT	960
GCTAGTTGGC AATAGATAAC TCAGCGTAGA TAGTATTGCA AAAAGGGAG GAAATACACA	1020
55 ACAATAATAA ATGTAAAAAC CTGCTATTCA ACATGCAGTT TTATTCGAR GCCAAAAATC	1080
TAGAGCTTTC CCAAGATCCT GTTGCCTTAG GCACATNCAC ACTTCAACAG TGCACACTAT	1140
60 CCAACAGTGC ACACATTCA ACAGTGCACA CTATTCAAAA GCGTAGACTA TTTTTTGCA	1200

TGTTCAAGAT	ATTTGTTTG	GTCTTATGTG	TGTGTGAGAG	AGAGAGATTG	CTTTGACATT	1260	
AAGGAGCATC	AATGAGAAAA	GATGATGAGG	CAGGAATTAA	TAAAGAAATG	AAGTCGTGTG	1320	
5	TGTTGGTTG	CCTGTCAGAG	GGCACACAAT	TTCATAAACCA	CCATGCCTGG	ACAATTTGAT	1380
	ATTAATATTT	AACACCTCTG	CATCTTTTC	TTAAAAAAGA	ATATGGCCA	GATACAGTGG	1440
10	CTCACATTTG	TAATCCCAGC	ACTTTGGGA	GCCAAGTTAG	CAGAATCCCT	TGAGCACAGG	1500
	AATCTGAAAC	CAGCTTGGGC	AACATAGTGA	GATCCCACAT	NTACAAAAAA	CTTAAAAAATT	1560
	AGCCAGGCAT	GATGGCACAT	TCCTGTAGTC	CTAGCTACTC	AGGAGGCTAA	GGTAGGAGGA	1620
15	TTGCCTGAGC	CCAGGAGTTC	AAGGCTGCAG	TGAGCTAAGN	ACGTGCCAGT	ACACTCCAGC	1680
	CTGAGCCACA	AAGTGAGACC	CTGCTCGCA	AAAAAAAAAN	TTAAAAAGTC	GGGGGGGGGC	1740
20	CCGGTACCCA	AATGCCCGGA	TATGATCGTA	AACAATC			1777

(2) INFORMATION FOR SEQ ID NO: 139:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 643 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

TTTTTTTTTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTTGGG	AATGAGAAAA	TAACTTTATT	60	
35	TTCATTGTGG	GGAGGGGGCC	GATGCCAGC	CTCAGAACCTT	CTGGAACCTGC	TTCTTGGTGC	120
	CGGCAGCCCT	GGTGACCTTG	AGCACCGTGA	AGCGCACTGT	CTTGCTCAGA	GGCCGGCACT	180
40	CGCCCCACTGT	GACGATGTCA	CCGATCTGGA	CGTCCCTGAA	GCAGGGGAC	AGGTGTACAG	240
	ACATGTTCTT	GTGGCGCTTC	TCGAAGCGGT	TGTACTTGCG	GATGTAGTGC	AGATAGTCTC	300
	GGCGGATGAC	AATGGTCCTC	TGCATCTTCA	TCTTGGGTCA	CCACGCCAGA	GAGGATCCGC	360
45	CCTCGAATGG	ACACATTACC	AGTGAAGGGG	CATTTCTTGT	CAATGTAGGT	GCCCCCTCAAT	420
	AGCCTCCTTG	GGGTGTCTTT	GAAGCCCAGA	CCGATGTCT	TGTTAGTAAC	CCGGGGGAGC	480
50	TTCTCCTTGC	CAGTTCTCC	CAGCAGGACC	CTCTTCTTGT	TTTGAAGAT	GGTCGGCTGC	540
	TTTTGGTAGG	CACGCTCAGT	CTGAATGTCC	GCCATCTTCT	CGTGCCGMAY	TCCTGCAGCC	600
55	CGGGGGATCC	ACTAGTTCTA	GAGCGGCCGC	ACCGCGGTGG	AGC		643

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1220 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

10	GGCACGAGGA TGATAGACCT ACTGGAGGAA TACATGGTTT ACAGGAAGCA TACCTACATR	60
15	AGGCCTGATG GCTCATCCAA GATCTCGGAG AGGGGAGACA TGGTGTGCTGA TTTTCAGAAC	120
20	AGGAATGACA TCTTTGTGTT CCTGTTAACG ACACGAGCTG GAGGACTGGG TATCAATCTC	180
25	ACTGCTGMAG ACACAGTGCA TTTTCTATGA TAGCGACTGG AACCCCACIG TGGACCAGCA	240
30	GCCCCATGGAC AGGGCCCACC GCTTAGGGCA GACAAAGCAG GTTACTGTGT ACCGGCTCAT	300
35	CTGTAAAGGC ACCATTGAAG AACGCATTCT GCAGGAGGCC AAGGAGAAGA GTGAGATTCA	360
40	CGGGATGGTG ATTTCACCGTG GGAACCTCAA ACCAGATAAC TTGAAACCCA AAGAGGTGGT	420
45	TAGTCTTCTT CTAGACGACG AAGAGTTGGA GAAGAAACGT ATGTACTCTA AACCTCTATA	480
50	CACTCCCCCTC ACGTATCTGA GAATGGAAGA GGTACTTGGS TGTGTGCCAA GGGTTAGGCA	540
55	AAGCCAGAGG CTGTATTTAG GGAAAGTATT TTTGTGCTCA TATTTTATAT AAAAACCAA	600
60	ACAAGAATGT GTTGTAGGC CAGGCGTGGT CGCTCGGCC TCTAGTCTCA GCATTTCGGG	660
65	ARGCCAAAGT GGGCAGATCA CCTGARGTCG GGARTTGAG TTTGARACCA GCCTGGCCMA	720
70	CGTTGTGAAA CCCCACCTCT ACTARGARTA CSGAAAATTG GTTGGGCATG GTGGGGGCA	780
75	CCTGTAATTG CAGCACTTG GGAGGCTGGG GCAGAANAAT TGCTTGAGCC CAGGAGGTGG	840
80	AGATTGCGGT GAGCCGAGAT YGTGCCATTG CAMTCCAGCC SGGCAATAA GAGTGAAAYT	900
85	CCATCTTTTA AAAACAAACA AAAACAAAAA ACACAAAGACG GCTCACACCT GTAATCCCAG	960
90	CACTTTGGGA RGCCGARGCA GGTGGATCAC GARGTCAGGA GTTCCAAGAC TAGCCTGGCC	1020
95	AACCTGGTGA AGCCCCGTCT CTACTAAAAA TACMAATATT AGTCGGCGT GGTGGTGGGC	1080
100	ACGTGTAATC CCAGCTACTC GGGAGGCTGA GGCAGGGAGAA TCCCTTGAAG CTAGGAGGCA	1140
105	GAGGTTGCAG TGAGCCAGGA TCGTGCCTATT GCACTCCAGC CTGGACAACA AGAGCAAGAT	1200
110	TCCATCTCAA AAAAAAAAAA	1220

(2) INFORMATION FOR SEQ ID NO: 141:

55	(i) SEQUENCE CHARACTERISTICS:
60	(A) LENGTH: 721 base pairs
65	(B) TYPE: nucleic acid
70	(C) STRANDEDNESS: double
75	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

	AATCGGCAC GAGCCAGGT AGCCGGAAGG GCAGCTCTCC AGGCCTGCC CACCCCACAG	60
5	GGGGCTCCCT ATGCACACCG GGGCGTCTCC TTGTGGCCAT AGAAACGGAA CTGGCTCTTT	120
	TCAACAGTGC TGCAAGAGGA TGGTTATTAA ACGCTGGCCC CCAAGGAGGA AAGGCACAGA	180
10	CYTTCCTCCC TCCTGGAACA TCCAAGGGCA CTGGATCCTC TGTCCTCCCTC TGAGATGGGG	240
	TGCCACTCCA GCAAGAGCAC CACGGTGGCA GCTGAGTCCC AGAACCTTGA AGAACAGYGC	300
15	GAGGGAAAGAG AGCCAGGTCT GGAGACCGGC ACCCAGGCAG CAGACTGCAA GGATGCCCG	360
	CTGAAGGATG GAACCCCTGA GCCAAAGAGC TGAAATGCCT CTCTCCAGAG TCGGACCCCTC	420
	ACCTCYTTCC TGGAACTGCC TTTGGCCCCA GAACCATGAG ACAATCCCCA CCCTGAGAAG	480
20	CTCCGATCAC TGGGAGGAGA GAGAAAGCCT CCAGCTTTGG GATTCAAGGCT TCAGAAGTTT	540
	TTAGCAGCCT TTGCTCATTG GAGAGGTGGG GAAAGGATAA AGTTCCTATA AGGAAATCCC	600
25	TAATTTCCCC CAGCTCCCTCC CCNCCNGAAG AAGGAACNAAG AGAAAGTTCC TTCCACACGT	660
	TTTGTGGAA ACTTTTCCCT TGCCAACCTT CCTTGGATTG CCAGAACAAA GCCCTCCAGA	720
	A	721
30		

(2) INFORMATION FOR SEQ ID NO: 142:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1468 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

	ATGAATTAAT GTTTATAAAAT GACTGTACTG AATTAAAAAC CGTACAGTTT CATTGCAATT	60
45	TTGACATTAC TTTATTATAC ATTTCGCATT TAAAAGGCTG CACCAGTTGG CTTTTCTCT	120
	GTTCATTCCT CAAAATATAG AGATTCCTGTG ATTATTTGC CCTGTTTATG GATTAAGAAAG	180
	AAAATTCTAA TATAAAGCAT TTCAATAGGA TGCATAGGTA TATTACGTTT TTTAAATGCT	240
50	TTAGATCTGT GATTCTTGAC TTACTATTAA TTTTATCCCC TTTAAGTCAG GGATGCTTTA	300
	TTCTATTTTA AAGCACTTAT GAGTTACATG TTGTAATCAA GTTGCACAA TATATTTATC	360
55	TATATGAGGA ACCCATAAAAT GAATAGCTAA TTTTTAAAT GCCATTAAAA TGCATGAAAT	420
	KCTTATTAAA ACCTTACTAT ACTATTTCTT CAAGGCAAGT AAATTGACCA TGRGRAAAGR	480
60	ACACAGTTAT TAAACACTGT TGACAGGAAA ATTCTCCCTG ATAACATAGG ACAATTAAATG	540

	GAAAAAAA TTCTCATTAT TTGCAAAGAA TGAACAAGTT AATGAACAAA CAAACTAGAT	600
	TTGGTATGTT TTCAGCTTT GTATCATGTT TAATTGTTA ATTTGGTTGA AAAACTGCAG	660
5	TTGAGAAATC AGATAGCAAT ATAGACATTC ACAGCAGCTC TGTGGATACC ATGTAATTGT	720
	CAGGTAAATT CAGAACATTC AAAATTATTC AGTGCAGCCC TCATAGTATC ATACTTGAAG	780
10	AAATTGATTA CAGTCCACT AAATTGTTGA AGATAAATTA TTTTAAAGG TTATGAAAAC	840
	TAAGTTATAT TAATTCAATAT GTTGATTT TAAATCCCAC CTCCCAAGC TATCCAATT	900
	NCTGACTTTG AAAATAACCA TGAGAGATGC CACATTCTC TCTGGAAAC TACCACTCAA	960
15	AGAATAATTG TTAAAAATTA AGCTTTAGG TATTAGAAGC TGTTATAAAG TATAAAATTA	1020
	AGATATAAGC AGATCACATG TAAATCAATC CTAAAGCACA AGAAAAGAAT GTGCCTTGAT	1080
20	GTACATATAT TACTAAGTTG CCTCTCCCAG TTTACTTTAA AAATGGCTTT AAGGATAAAG	1140
	AATAATGTT ATAGCTGTGC ATGCATTATA TATTGTCATT TCCAAATTTC CCATTGTTT	1200
	AACAGCTGTG TGGCTGACTT TCAATTAA GACGTGAATT GACATACAGC CCATAACTTT	1260
25	ATAATGGCTG CTCATTATC TTATCTTCA GTTAGTGGAA AAACATTCA ACCTGACTAA	1320
	AATTTGGAAT TGTGTCTTT ATGTTCCATC CTCTGTTGTT ACTAGATTAA GTTTAAAAAT	1380
30	TGTGTATGAC CATTAATGTA TGTACATAAAC ATGTAATTA AAGATGTTGA ATCTTGTGA	1440
	AAAGCAWRAA AAAAAAAA AAACTCGA	1468

35

(2) INFORMATION FOR SEQ ID NO: 143:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 300 base pairs	
40	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:	
45	TGAATTTCCTT GCCAAACTTA GTAACTCTGT TAAATATTG GAGGATTTAA AGAACATCCC	60
	AGTTTGAATT CATTCAAAAC TTTTAAATT TTTTGTACT ATGTTGGTT TTATTTCT	120
50	TCTGTTAAC TTTGTATTTC RCTTATGCTC TCGTACATTG AGTACTTTTA TTCCAAAATC	180
	AGTGGGTTTT CTCTACTGGA AATTTCAAT AAACCTGTCA TTATTGCTTA CTTTGATTA	240
55	AAAAAAAAA AAAAAAAA AAACCCNAG GGGGGGGCCG CGTNCCTAAT CCCCCCCCCAAA	300

60

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2243 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

	TGCCCTCCCTT CCTGCAGATT GTGGACAGTA GTTCCTCAGC CTGCACCCCTG GATTCCCTCT	60
10	TCCCCCTTCCT AGCTCCATGG GACTCGCCCC AAGACTGTGG CTTCAAGGAC CACCAGCCCC	120
	TTACTCTTCA AGCCCTGACT GTGGAGGTGG TAGATGCCCTC TGATCCTCAG TATTCTCTCT	180
15	GGCAATGPTC CACGGCTCTC CCTTCCTGGG AGCTGGCTCC ATAACCTGAT TTTCCCCAAA	240
	CGTGTGCAA TCCCTGCTGC CCCTTAGCCA CCCAGGGTCT TGTGTGGTA TGAGTGTAGA	300
20	GGATGGGGT ATGCCAGGCC TGGGCCGTCC CAGGCAGGCC CGCTGGACCC TGATGCTACT	360
	CCTATCCACT GCCATGTACG GTGCCCATGC CCCATTGCTG GCACITGTGCC ATGTGGACGG	420
	CCGAGTGCCC TTTCGGCCCT CCTCAGCCGT GCTGCTGACT GAGCTGACCA AGCTACTGTT	480
25	ATGCCCTTC TCCCTCTGG TAGGCTGGCA ACCATGCCCT CAGGGGCCCT CACCCCTGGCG	540
	CCAGGCTGCT CCCTTCGGCAC TATCAGCCCT GCTCTATGGC GCTAACAAACA ACCTGGTGTAT	600
	CTATCTTCAG CGTTACATGG ACCCCAGCAC CTACCAAGGTG CTGAGTAATC TCAAGATTGG	660
30	AAGCACAGCT GTGCTCTACT GCCTCTGCCT CCGGCACCGC CTCTCTGTGC GTCAGGGGT	720
	ACCGCTGCTG CTGCTGATGG CTGCGGGAGC CTGCTATGCA GCAGGGGGCC TTCAAGTTCC	780
35	CGGAAACACC CTTCCCAGTC CCCCTCCAGC AGCTGCTGCC AGCCCCATGC CCCCTGCATAT	840
	CACTCCGCTA GGCTCGCTGC TCCCTCATTC GACTGCTCTC ATCTCAGGCT TGTCGTCAGT	900
	GTACACAGAG CTGCTCATGA AGCGACAGNG GCTGCCCTG GCACITTCAGA ACCTCTTCCT	960
40	CTACACTTTT GGTGTGCTTC TGAATCTAGG TCTGCAATGCT GGCGGCGGCT CTGGCCCAGG	1020
	SCTCCTGGAA GGTTTCTCAG GATGGGCAGC ACTCGTGGTG CTGAGCCAGG CACTAAATGG	1080
45	ACTGCTCATG TCTGCTGTCA TGAAGCATGG CAGCAGCAGC ACACCCCTCT TTGTGGTGTG	1140
	CTGCTCGCTG GTGGTCAACG CCGTGCTCTC AGCAGTCCTG CTACGGCTGC AGCTCACAGC	1200
50	CGCCTTCCTC CTGGCCACAT TGCTCATTCG CCTGGCCATG CGCTCTGACT ATGGCAGCCG	1260
	CTAGTCCCTG ACAACTTCCA CCCTGATTCC GGACCCCTGTA GATTGGGCC CACCACCAGA	1320
	TCCCCCTCCC AGGCCTTCCT CCCCTCTCCC TCAGCAGGCC TGTAACAAGT GCCTTGTGAG	1380
55	AAAAGCTGGA GAAGTGAGGG CAGCCAGGTT ATTCTCTGGA GGTTGGTGGTA TGAAGGGTAA	1440
	CCCCTAGGAG ATGTGAAGTG TGGGTTGGT TAAGGAAATG CTTACCATCC CCCACCCCCA	1500
60	ACCAAGTTCT TCCAGACTAA AGAATTAAGG TAACATCAAT ACCTAGGCCT GAGAAATAAC	1560

	CCCATCCTTG TTGGGCAGCT CCCTGCTTGT TCCTGCATGA ACAGAGTTGA TGAAAGTGGG	1620
	GTGTGGCAA CAAGTGGCTT TCCTTGCTA CTTTAGTCAC CCAGCAGAGC CACTGGAGCT	1680
5	GGCTAGTCCA GCCCAGCCAT GGTGCATGAC TCTTCCATAA GGGATCCTCA CCCTTCCACT	1740
	TTCATGCAAG AAGGCCAGT TGCCACAGAT TATACAACCA TTACCCAAAC CACTCTGACA	1800
10	GTCTCCTCCA GTTCCAGCAA TGCCTAGAGA CATGCTCCCT GCCCTCTCCA CAGTGCTGCT	1860
	CCCCACACCT AGCCCTTGT CTGGAAACCC CAGAGAGGGC TGGGCTTGAC TCATCTCAGG	1920
	GAATGTAGCC CCTGGGCCT GGCTTAAGCC GACACTCCTG ACCTCTCTGT TCACCCCTGAG	1980
15	GGCTGTCTTG AAGCCCGCTA CCCACTCTGA GGCTCCTAGG AGGTACCATG CTTCCCAC	2040
	TGGGGCCTGC CCCTGCCTAG CAGTCTCCCA GCTCCAAACA GCCTGGGGAA GCTCTGCACA	2100
20	GAGTGACCTG AGACCAGGTA CAGGAAACCT GTAGCTCAAT CAGTGCTCT WTAACTGCAT	2160
	AAGCAATAAG ATCTTAATAA AGTCTCTAG GCTGTAGGGT GGTTCCCTACA ACCACAGCCA	2220
	AAAAAAAAAAA AAAAAAAACTC GAG	2243
25		

(2) INFORMATION FOR SEQ ID NO: 145:

30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1082 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:	
	GCCAAGCTCT AATACGACTC ACTATAGGGA AAGCTGGTAC GCCTGCAGKT ACCGGTTCCG	60
40	GGAATTCCCG GGTGACCCCA CGCGTCCGCT TCCGTGTGTC AAAATCCTCA CCTCCCTTCAT	120
	AACCATCTCC CACAATTAAAT TCTTGACTAT ATAAATTTAT GGTTGATAA TATTATCAAT	180
45	TTGTAATCAA TTGAGATTTTC TTAGTGCTT GCTTTCTGT GACTCAACTG CCCAGACACC	240
	TCATTGTACT TGAAAACCTGG AACANCTTGG GAATGCCATG GGGTTTGATA ATCTGCCAGG	300
	GACATGAAGA GGCTCAGCTT CCTGGGACCA TGACTTTGGC TCAGCTGATC CTGNACATGG	360
50	GAGAACAAACC ACATTTTCTT TTGTGTGTGC TTCTAGCAGC TGTTGGGAG GACCKTGACC	420
	CAAYAGTGTGTT CCCATGCTGT TTCTTGTGAA ATGCTCTCGG CTATGTAGCA GCTTTTGATT	480
55	CCCTGCATAC CCTAGGCTGC TGCCCCTATC CTGTCCTTG TTTATAACAT TGAGAGGTTT	540
	TCTAGGGCAC ATACTGAGTG AGAGCAGTGT TGAGAAGTCG GGGAAAATGG TGACTACTTT	600
	TAGAGCAAGG CTGGGCATCA GCACCTGTCC AGCTCTACTT GTGTGATGTT TCAGGAAC	660
60	AGCCCCCTTTT TCTGCCCTAGG ATAAGGAGCT GAAAGATTAA CTTGGATCTY CTAATGGTCC	720

AAATCTTTTG	GTCACAATAA	AGAGTCTCCA	AATTAGAGAC	TGCATGTTAG	TTCTGGATGG	780	
5	ATTTGGTGGC	CTGACATGAT	ACCCCTGCCAG	CTGTGAGGGG	ACCCCGTTTT	TGAGATGCAT	840
	GGCCAAGCTC	TCTGCAAATG	GAAATGCTTA	CACTGGGTGT	TGGGGATGTT	TGCTACCTCC	900
	TGCTATTTTT	GTGGTTTGG	TTCTCCACT	ATCGTAGGAC	CCCTGGCCAG	CATTTGGCT	960
10	TGTCATGTCA	GCCCCATTGA	CTACCTCTC	ATGCTCTGAG	GTACTACTGC	CTCTGCAGCA	1020
	CAAATTCTA	TTTCTGTCAA	TAAAAGGAGA	TGAAAATAAA	AAANAAAAAA	AAAAAAACTCG	1080
15	NG						1082

(2) INFORMATION FOR SEQ ID NO: 146:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4313 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

30	CAAGCTGGTT	TGAAACTAGG	GGTCGGGCTC	GGCCGTCGTC	GTTGTTGTC	GCCGCATCCC	60
	CGCTTCCGGG	TTAGGCCGTT	CCTGCCGCC	CCCTCCCTCTC	CTCCCTTCGG	ACCCCATAGAT	120
	CTCAGGCTCG	GCTCCCCGCC	CCCGCGAGCC	CACTGTTGAC	CGGGCCCGTA	CTGGGGCCCC	180
35	GTGGCCACCA	TGTCCCTGCA	CGGCAAACGG	AAGGAGATCT	ACAAGTATGA	AGCGCCCTGG	240
	ACAGTCTACG	CGATGAACTG	GAGTGTGCGG	CCCGATAAGC	GCTTTCGCTT	GGCCCTGGGC	300
40	AGCTTCGTGG	AGGAGTACAA	CAACAAGGTT	CAGCTTGTG	GTTTAGATGA	GGAGAGTTCA	360
	GAGTTTATTT	GCAGAACAC	CTTGACCAC	CCATACCCA	CCACAAAGCT	CATGTGGATC	420
	CCTGACACAA	AAGCGTCTA	TCCAGACCTA	CTGGCAACAA	GCGGTGACTA	TCTCCGTGTG	480
45	TGGAGGGTTG	GTGAAACAGA	GACCAGGCTG	GAGTGTGTTGC	TAAACAAATAA	TAAGAACTCT	540
	GATTTCTGTG	CTCCCCCTGAC	CTCCTTGTGAC	TGGAATGAGG	TGGATCCCTA	TCTTTTAGGT	600
50	ACCTCAAGCA	TTGATACGAC	ATGCACCAC	TGGGGCTGG	AGACAGGGCA	GGTGTAGGG	660
	CGAGTGAATC	TCGIGTCTGG	CCACGTGAAG	ACCCAGCTGA	TCGCCCATGA	CAAAGAGGTC	720
	TATGATATTG	CATTTAGCCG	GGCCGGGGGT	GGCAGGGACA	TGTTTGCTC	TGTGGGTGCT	780
55	GATGGCTCGG	TGCGGATGTT	TGACCTCCGC	CATCTAGAAC	ACAGCACCAC	CATTTACGAA	840
	GACCCACAGC	ATCACCCACT	GCTTCGCCCTC	TGCTGGAACA	AGCAGGACCC	TAACTACCTG	900
60	GCCACCACATGG	CCATGGATGG	AATGGAGGTG	GTGATTCTAG	ATGTCCGGGT	TCCTGCACAC	960

	CTGTSGCCAG GTTAAACAAAC CATCGAGCAT GTGTCAATGG CATTGCTTGG GCCCCACATT	1020
	CATCCTGCCA CATCTGCACT GCACCGGATG ACCACCAGGC TCTCATCTGG GACATCCAGC	1080
5	AAATGCCCG AGCCATTGAG GACCCTATCC TGGCCTACAC AGCTGNAAGG WGAGATCAAC	1140
	AATGTGCAGT GGGCATCAAC TCAGCCGAA YTGTGCCAT CTGCTACAAC AACTGCCCTGG	1200
10	AGATACTCAG AGTGTAGTGT TGGTGGCGCT GTGCCACGA GGCAGGGCT TTTGTATTC	1260
	CTGCCCTTGC CCCACCCCCA AAGTAAGAAG AAACATGTT CCAGTGGCCA GTATGTCCTT	1320
	CATTGCTTTG CACCCACTGT TACCAAGAAGC TGCTCTAGGA GTTCCTGGCC AGTCACCCCCA	1380
15	TGGCCCTCTG TGGCAGACTC AGTGCTGTGT GGCGCCTCCT CAGCCCAGGG CTGAGTTTA	1440
	AGATTTCTC TCCTTTCTC TTCTCCTTTG GTTCCTCAAT TAAAAAAATGT GTGTATATT	1500
	GTTTGTCAAG CGTTGTGTTG AGGAGCAGTT CACGCAGTGG CTGIGCTAT TCCTCTGCC	1560
20	AGGTGTCTCT GTTTGCTGCC CAAKGYWKKT TTTCATGTC CGTCCATGTC CATGTTGTC	1620
	TTAGCACTWA CGTGGGAACA AATACCAATT TGTCTTTCT CCTAGTATCA GTGTGTTAA	1680
25	CAAATTTAA CTTTGTATAT TTGTTATCTA TCAGGCTAAT TTTTTATGA AAAGAATT	1740
	ACTCTCCCTGC TTCAATTCTT TGTCTTATAG TCCTCCCTCT TTGCACCTTC TTCTCTCCC	1800
	TCAGTGCCTG GAGCTGGTAC TGGGCCCCCTG GCCCCATGAG CAGTTGCTT TCTTGAGICA	1860
30	CTGCCCTGTGT AGTACATACC TGACCGGGAG TCCAAACAC CTTGGTGTC TGAAAGTCCAC	1920
	TGACTCATCA CACCTTTCTT AGCCTGGCTC CTCTCAAGGG CATTCTGGC TTGTAAACAG	1980
35	ACATAGGAAG CCTCTGTTA CCCTGAAGCA CCACTGTCCA GCCCATTGGT TCCCACGGC	2040
	AGCATGGTAG AGCTGAGAGA AACAGGCTCT CAGGGTACCT GACTTGAGGG GAATCGTTTC	2100
	ATGAAGCTGA ACTTCAAGCA TATTTCCAGT ACATTCTTTC AGAGTCTGTT TTTCCATCCA	2160
40	AATATAAGCC CCAGGCCATT CCACTTAGTG TCTTTCAAT GATAGGCAAG AATGATAATCT	2220
	GAGTTGAAC TCGGTGCTTC TGTGTGTTGA GTTTACTGIG CCTGGTGGTA TATTGGGCAT	2280
45	TCTTTGGATT GAGTGTCTG AGGTGAGAGA GTCTTCCGA GGCACTCTGT CTGTGCTTCC	2340
	AACCCCTGAAC AAGACCTTAC ATGAGAGATG GACTGATGGA CTGGGCAAT CCTGGGCTGT	2400
	CAAGTGGATA GATAGTTAAA AAGCATTATA CTGTGGTAA TGAAAAGGGA GGAAAAAAA	2460
50	AGAAGGAAA GGAATTATAG ACCCCCAGGG TCAGCCAGTT AAGAGCTCTA CCCACACCTG	2520
	TCAACCCCTC TCTCCCCAG TTTAGGTTCT GAGCACTTATT GGACTTGTAG CCTGCAGTTG	2580
55	TCTTTTGACT TGCAGGCCGC AGTGTCTTTC TGTTATGTGA ATGAGTTCCA TGGAGGGCA	2640
	TATGTGTGAT TCCACCGTTA GATGAGCCCT TGGGGCAGGC AGTTGGGAT GTGCTCTTGG	2700
60	GGGAAAGTTG GCTGTTCCT TGGCCTCTGC TCCTACCGA AGTTTTTAAG TCCCTCTGAA	2760

	TTGCTCATCT GAGATTAGTA GAGTAGCAGG CCTGAAGGAT GATGGTTTG TCCTCTTGG	2820
	TTCTCACCTG CTTGAGAAGT AAAACAGTAA CTTTGTCTT CTGGGCCCTT AAGCTTTTT	2880
5	GGTTAACGTCT TCCTTTCAAG AAGTAGATGT CATTATATGC CAAAAGTCTA GCTCTTTGCT	2940
	TTACCATACA GGGACCTGTC CCAAAGAAAA AGGCTCTTT TTTAGCCAGC ATATTTCCCC	3000
10	TTCTACCCCTT TTACTTTGTT GTTCTGATTT TAGGACTCTG GCTGCCATG TGCTTGTGGT	3060
	TGCCCTCTCC GCATTTGCCA CTGGATTGTC ACTGCATCGT TTGGAGATAC AAAGCGAGCA	3120
	GTTCTTGGTC AGAACCCCTCC TCTGCTTTTC ATTGTGTTTG ATAATGGTTA CTGGGTCCCT	3180
15	CTCTCAAGGG TAGCAAGGCC AAGCTGATGG CTGCTTGT TT AGGAGGCCAT CAGTCCTTC	3240
	CTGTGGAGAA GGGTCTGAAA TGGAAGTCAG TGGTAGAAGG GGCTGGTCTG CTGGGCAGGG	3300
20	CTTACATCCA CTGAGTTCTA AGATTCTTT CCTGATCTGC ACCTACGCC CAGTCGTATG	3360
	GTGGAATTG TCAGCTGGAA CTCAGAAACA ACAACTTGAA AAAAAAATAA TAATTAGAAC	3420
	ATATTTGCAT AAGATAGCTA TTTACTCTGG AAACCAACAA CTTTGAGAT TTCCCTTGCC	3480
25	CTGTGGACGC CCAGCTCTG TCATCCCTCC TTAGGTCTG CAGTACAGTC TTCCCTGAA	3540
	TGCCACCGGG GACCCAGGG GACTCCACCC CCCTAAGCAA GCACACACAT ACTCACAGTT	3600
30	GATGAGTTGC TGGCTTTGA GTCCCAGCTC TCTTACCTTC CCTTTACTCC ACCAGCCGA	3660
	CGACCCATGA CTGAGGAGGG GATTTCTACA GTCTCAGGAT TTAGAAAGTC TGTAAGCCAT	3720
	CCATGCTCCA GAAAGCACCG ATCTGTTGTA GTTGCAAAAA CAACTCTGTA ATTTGTTGAG	3780
35	GTTCTAAAC TGACAGCCAG CGAGACTGGG TGGGAGGCC TGGATCTGTT CTCCCTGACT	3840
	GCGGGAGGAG CAGCCACTAG GACTTTAGCA GGAAGCCAC ATGGAGGCTC CGCCAGGCTG	3900
40	TGGCCACCT GGTGATGGCC CTTTGCTCC TGGCAGCCTG AGGCACAGCT GCCTGTATG	3960
	TCCTCATCTG TTCTGACTGA AGGATGGAGG TGCTGAATAA ATTAGGCCTC AGGCNTCTAC	4020
	CACCAAGAG CTGGAGAATG GGTCCACGTC ATTCAAGGAC CTGAATTGTT TATGCTCAGG	4080
45	AGCATTGGAA TCCCTCTCTT CCAGGGAGGA ATTAGGCCTG AAGGTTAGGA CTTGAAGAGG	4140
	GAAGGTATTT AATAACTGGG CGAGGATGGG TGTGGTGGCT CACACCTGTA ATCCCAGCAT	4200
	TTTGGGAGGC TGAGGTGGCC AGATCCCAAG GTCAAGAAGAT CGAGACCATC CTGGCTAACAA	4260
50	TGGTGAAACC CCATCTCTAC TAAAAATACA AAATTAATT GGCCGGGGGT GAA	4313

55

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1183 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5	GGCAGAGCCT CAAGCTGACT TGGATTATGT GGTCCTCAA ATCTACCGAC ACATGCAGGA	60
	GGAGTTCCGG GGCGGGTAG AGAGGACCAA ATCTCAGGGT CCCCTGACTG TGGCTGCTTA	120
10	TCAKWYGGGG AGTGTCTACT CAGCTGCTAT GGTCACAGCC CTCACCCCTGT TGGCTTCCC	180
	ACTTCTGCTG TTGCTATGCC AGCGCATCAG CCTTGTGTTTC CTGCTCTGT TTCTGCAGAG	240
15	CTTCCCTCTC CTACATCTGC TTGCTGCTGG GATAACCGTC ACCACCCCTG GTCCCTTTAC	300
	TGTGCCATGG CAGGCAGTCT CGGCCTGGGC CCTCATGGCC ACACAGACCT TCTACTCCAC	360
	AGGCCACCAAG CCTGTCTTTC CAGCCATCCA TTGGCATGCA GCCTTGTGG GATTCCCAGA	420
20	GGGTCATGGC TCCTGTACTT GGCTGCCTGC TTIGCTAGTG GGAGCCAACA CCTTTGCCTC	480
	CCACCTCCCTC TTGCACTAG GGTGCCACT GCTCCTGCTC TGGCTTTCC TGTGTGAGAG	540
	TCAAGGGCTG CGGAAGAGAC AGCAGCCCCC AGGGAATGAA GCTGATGCCA GAGTCAGACC	600
25	CGAGGAGGAA GAGGAGCCAC TGATGGAGAT GGGGCTCCGG GATGCCCTC AGCACTTCTA	660
	TGGCACTG CTGCAGCTGG GCCTCAAGTA CCTCTTATC CTTGGTATTIC AGATTCTGGC	720
30	CTGTGCCTTG GCAGCCTCCA TCCTTCGAG GCATCTCATG GTCTGGAAAG TGTTCGCC	780
	TAAGTTTATA TTGAGAGCTG TGGCTTCAT TGTGAGCACC GTGGGACTTC TCCGGGCAT	840
35	AGCTTTGGTG ATGAGAGTGG ATGGTGCTGT GAGCTCTGG TTCAGGCAGC TATTTCTGGC	900
	CCAGCAGAGG TAGCCTAGTC TGTGATTACT GGCACCTGGC TACAGAGAGT GCTGGAGAAC	960
	AGTGTAGCCT GGCTGTACA GGTACTGGAT GATCTGCAAG ACAGGCTCAG CCATACTCTT	1020
40	ACTATCATGC AGCCAGGGGC CGCTGACATC TANGACTTCA TTATTCTWTR ATTCAAGGACC	1080
	ACAGTGGAGT ATGATCCCTA ACTCCTGATT TGGATGCCATC TGAGGGACAA GGGGGKCGGT	1140
	STCCGAAGTG GAATAAAATA GGCGGGCGTG GTGACTTGCA CCT	1183
45		

(2) INFORMATION FOR SEQ ID NO: 148:

50	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 734 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
55	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

60	GAATTCCGCA GAGTGAAGCA TTAGAATGAT TCCAACACTG CTCTTCTGCA CCATGAGACC	60
----	-------------------------------------------------------------------	----

	AACCCAGGGC AAGATCCCAT CCCATCACAT CAGCCTACCT CCCTCCCTGGC TGCTGGCCAK	120
	GATGTCGCCA GCATTACCTT CCACTGCCIT TCTCCCTGGG AAGCAGCACA GCTGAGACTG	180
5	GGCACCAGGC CACCTCTGTT GGGACCCACA GGAAAGAGTG TGGCAGCAAC TGCMGGCTG	240
	ACCTTCTAT CTTCTCTAGG CTCAGGTACT GCTCCCTCCAT GCCCATGGYT GGGCCGTGGG	300
10	GAGAAGAACG TCTCATACGC CTTCCCACTC CCTCTGGTT ATAGGACTTC ACTCCCTAGC	360
	CAACAGGAGA GGAGGCCTCC TGGGGTTTCC CCRROGCAGT AGGTCAAACG ACCTCATCAC	420
	AGTCTTCTT CCTCTTCAAG CGTTTCATGT TGAACACAGC TCTCTCCRCT CCCTTGATGAT	480
15	TTCTGAGGGT CACCACTGCC ARCCCTCAGGC AACATAGAGA GCCTCCCTGTT CTTTCTATGC	540
	TGGTCTGAC TGAGCCTAAA GTTGAGAAAA TGGGTGCCAA GGCCAGTGCC AGTGTCTTGG	600
20	GGCCCCTTTG GCTCTCCCTC ACTCTCTGAG GCTCCAGCTG GTCTGGGAC ATGCAGCCAG	660
	GACTGTGAGT CTGGGCASGT CCAAGGCTG CACCTTCAAG AAGTGAATA AATGTGGCCT	720
	TTGCTTCTAT TTAA	734
25		

(2) INFORMATION FOR SEQ ID NO: 149:

30	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1405 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:
----	--------------------------------------------

	GGCACAGTGG ACCCCAGACT CCCCTCCGC CTTTCTCTGC CTGGGGAGAC CCACTGTGTG	60
40	CATGGCATCA CTGACTCCCA TACCTCTGGC TATCAAAGGT TTCTGCCATG GCCACCCCTGG	120
	AAGSAAACCA GAGGGAGGTA GACAGGGAGA TCAGGTCCCT TCTACTCTGG TTCCTGCTCT	180
45	GTGAAATTGT CTCAGGCTGG CTGTGTCCAG ARGGTCCCTG GTTCTCTCAR GGATGCCAAA	240
	TCTACAAGAA TCTCTCCCTC TCCAGTTCT ATAACCTCTC CTTCTTTTG TCTCTTCTAGA	300
	CCTTGGAGTA GTAGCAGCCA GGTTCTTCT ATCTCTGGGT TAGTGCATTA TCTCTGGTGG	360
50	CTCCCTTACC CAGGACTTIG GGAATGGTCT TTTTGTATA CATTCTCCCTC AAATAATTCA	420
	ATTTTGAGTG TTCTGTATGT ATCCTGCTGG GAGGTGTTA TATACAAATC ACTGTGCCCG	480
	TTTAGCAGAG AAGGAGACTG AAGCTCAGGG AGGTAAAGTG TCTTCTCTA GGCGTATTG	540
55	TGGAGAAAGT GGCTGACTGG GGACTTGAAT GAGGTCCCTA GTTTCATGCT CGGAGGGCAA	600
	AGANGAATGT CCAATTGGCC TGAGATAAGC CTCTGGTAAA ATGTACTGTA CATAATAGGT	660
60	AATCAATAAA TGTGGCTGA TGACAAACAT GTTTCTTTG TTCATTAGTT ATAGTGATTA	720

400

	TGTTCTAAAT AACTCCMACA AGGAARTCAG CACATTGGA ATATCAWTAT CTTTCCATGA	780
5	TAATATCTTT CCMyGGAAAG AWAATGATAT TCCMAACTGG GAGTGTCCCW AGCARATCTG	840
	ANTCTGTGTA TTGGCCCTGG GGTGGGCCAG CCCCTTAGAC TCTATGGTCT CATTCTCTT	900
	GTTTACAAAA TTGAGATAAG GCCTTATTCT CTCCCCACCC CACCCATCCA TATTGTTTG	960
10	AGAATAAAAT GAGAGGATGT GTGTCAAGGG TGTATTTGG CAATAGTCTC TGAGCCATT	1020
	TCTGAGCACC TCCATACTGT TGACACTCAA GTAATATTTC ATCAGCATTG CATTCAAGGNT	1080
15	CCTCCCTTAA TGAGGTGTGC GATGTACAAG AGTYGTGAGG TGGCAAAGGA TGGGCTCCTG	1140
	AGGAAACACT TAGGAAACTG GGCTTTCTGC CATTAAAAGA GACAAACCTT TGTGGTGACC	1200
	TAATTAAAGT TTTTAAATT CAATTGGAA AGTTAGCAAG CTAGCTCCTK TCCAGGWAAA	1260
20	ATAAGGAGTC AGTGCATGAC CTAACCGGTC CGGGCTGCT TGCCATTCCA AACAACTGCA	1320
	GTAAGTTAT CACNTCTTT CAGGGACTGA GGTTCCAGG CACAGACTTG GATAAGGAAG	1380
	GATGTCCTAT GGGTCACAT TGATG	1405
25		

(2) INFORMATION FOR SEQ ID NO: 150:

30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2890 base pairs	
	(B) TYPE: nucleic acid	
35	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:	
40	TTATATGCTA CAGCTACAGT AATTCTTCT CCAAGCACAG AGGANCTTC CCAGGATCAG	60
	GGGGATCGCG CGTCACITGA TGCTGCTGAC AGTGGTGTG GGAGCTGGAC GTCATGCTCA	120
	AGTGGCTCCC ATGATAATAT ACAGACGATC CAGCACCAGA GAAGCTGGGA GCTCTTCCA	180
45	TTCGGGCATA CTCACTTGTA TTATTCAGGG GATCCTGCAG GTTTATGGGC ATCAAGCAGC	240
	CATATGGACC AAATTATGTT TTCTGATCAT AGCACAAAGT ATAACAGGCA AAATCAAAGT	300
	AGAGAGAGCC TTGAACAAGC CCAGTCCCGA GCAAGCTGGG CGTCTTCCAC AGGTTACTGG	360
50	GGAGAAGACT CAGAAGGTGA CACAGGCACA ATAAAGCGGA GGGGTGGAAA GGATGTTCC	420
	ATTGAAGCCG AAAGCACTAG CCTAACGTCT GTGACTACGG AAGAAACCAA GCCTGTCCCC	480
55	ATGCCTGCC ACATAGCTGT GGCATCAAGT ACTACAAAGG GGCTCATTGC ACGAAAGGAG	540
	GGCAGGTATC GAGAGCCCCC GCCCACCCCT CCCGGCTACA TTGGAATTCC CATTACTGAC	600
60	TTTCCAGAAG GGCACCTCCC TCCAGCCAGG AAACCGCCGG ACTACAACGT GGCCCTTCAG	660

	AGATCGCGGA TGGTCGCACG ATCCTCCGAC ACAGCTGGC CTTCATCCGT ACAGCAGCCA	720
	CATGGGCATC CCACCAGCAG CAGGCCCTGTG AACAAACCTC AGTGGCATAA AYCGAACGAG	780
5	TCTGACCCGC GCCTCGCCCC YTATCAGTCC CAAGGGTTT CCACCGAGGA GGATGAAGAT	840
	GAACAAGTTT CTGCTGTTTG AGGCACAGAC TTTTCTGGAA GCAGAGCGAG CCACCTGAAA	900
10	GGAGAGCACA AGAAGACGTC CTGAGCATTG GAGCCTTGGA ACTCACATTG TGAGGACGGT	960
	GGACCAAGTTT GCCTCCCTCC CTGCCCTAAA AGCAGCATGG CGSTTCTTCT CCCCTTCTTC	1020
	CTTTCCCCTT TGCATGTGAA ATACTGTGAA GAAATTGCCG TGGCACTTTT CAGACTTTGT	1080
15	TGCTTGAAAT GCACAGTGCA GCAATCITCG AGCTCCCCT GTGCTGCCT GCCACATCAC	1140
	ACAGTATCAT TCCAAATTCC AAGATCATCA CAACAAGATG ATTCACTCTG GCTGCACPTC	1200
20	TCAATGCCTG GAAGGATTTT TTTTAATCTT CCTTTTAGAT TTCAATCCAG TCCTAGCACT	1260
	TGATCTCATT GGGATAATGA GAAAAGCTAG CCATTGAAC ACTTGGGCC TTTAACCCAC	1320
	CAAGGAAGAC AAAGAAAAAC AATGAAATCC TTTGAGTACA GTGCTTGTC ACCTGTTTAC	1380
25	AATGICCTCC TTTTAAAAAA AAAAAATGA GTTTAAAGAT TTTGTTCAGA GAGTAAATAT	1440
	ATATCCATT AATGATTACA GTATTATTTT AAACCTTAAG TAGGGTTGCC AGCCTGGTT	1500
30	CTGAAAAACC AAATATGCCG GACAGGGTGT GGCCACACCA AGAAGACGGG AAGACCTGGC	1560
	TTGTGACCCCT GGCTTCCCAT GTCTTCTGG TCTCACCCGC GAAGTGCCT ATCCTGGAAG	1620
	TATGAAATGT TAGCCAATTA ATACCAAGAC ACCTCATCTG CTCCCTCCCC AGTGGATGGG	1680
35	GTTCTCTGT AAAACTGTTT GCACATGGCC AGGGGAGGGA ACTAGGACCC TTGTGTCCTG	1740
	TCTGAGCCTT ATGGAGGCAG GACGGTGTCA TTGGCGGATG TGTCTGCTC CATTGAGATG	1800
40	GATGGCAAAC CCCATTTTA AGTATATTTT CTTGATTTT TGTTAATTAA GAGGTGTAGG	1860
	TTTGTTTTT TGTTTTTTTG TTTTTTTTA AGAGAAACAT TTATAACTGG ATAGCATTGC	1920
	AGTGAAAGCA GCTTGGGATG TTGGAGCTAA TGCCAGCTGT TTATACTGCT CTTCAAGAC	1980
45	AGCCTCCCTT TATTGAATTG GCATTAGGGA ATAAACAAGC CTTAAACGT GATAAAAGAT	2040
	CAAAAACCTG GTTAGACATG CCAGCCATTG CAAGGCAGGT TAGTCACCAA AGACTAACCT	2100
50	CCAAGTGGCT TTATGGACGC TGCATATAGA GAAGGCCTAA GTGTAGCAAC CATCTGCTCA	2160
	CAGCTGCTAT TAACCCTATA ATGACTGAAA TGACCCCTCC ACTCTATTTT TGTGTTGTTT	2220
	TGCACAGACT CCGGAAAAGT GAAGGCCTGCC AATCTGAGTA GTACTCAAAT GTGAGGAACT	2280
55	GCTGGCTTG GATTTTTTT CCATTAATT CAGCTGATCA TATTGATCAG TAGATAAACG	2340
	TAAATAGCTT CAAATTTAA AAGTGGATT GCAGTGTGTTT TTCACTGTAT CAAACAATGT	2400
60	CAGTGCTTTA TTAAATAATT CTCCTCTGTA TCATGGCATT TGCTACTTG CTTATTACAT	2460

	TGTCAATTAT GCATTGTAA TTTTACATGT AATATGCATT ATTTGCCAGT TTTATTATAT	2520
	AGGCTATGGA CCTCATGTGC ATATAGAAAG ACAGAAATCT AGCTCTACCA CAAGTTGCAC	2580
5	AAATGTTATC TAAGCATTAA GTAATTGTAG AACATAGGAC TGCTAATCTC AGTCGCTCT	2640
	GIGATGTCAA GTGCAGAATG TACAATTAAC TGGTGATTTC CTCATACTTT TGATACTACT	2700
10	TGTACCTGTA TGTCTTTAG AAAGACATTG GTGGAGTCTG TATCCCTTTT GTATTTTAA	2760
	TACAATAATT GTACATATTG GTTATATTTT TGTGAAGAT GGTAGAAATG TACTATGTTT	2820
	ATGCTTCTAC ATCCAGTTTG TACAAGCTGG AAAATAAATA AATATAACAT AAAAAAAA	2880
15	AAAAAAAAAA	2890

20 (2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2399 base pairs
- (B) TYPE: nucleic acid
- 25 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

30	GAACTTTCC ATCTGGAAA CGGAAACTC CATCCCCATT AAACCAACTC CCCCTTTGG	60
	TTTCCCCCCC AGNGGAATAG AATTGGACN CCCATATAAA TCCAGGAAAC CACCTAAATT	120
35	CTTTAGTNGT TIGTGTGTC AAGATCTAAG GTCATGGTAA ACATTAAGTT CTTAAATTT	180
	TTGGGAGGGA CCAGTGCACC TCTCCCTCTG AATIGTCNC CAATTTAAA TTGGAGTAAG	240
	GTTTTAAAAT GTCTNATTCC ATTGGAAGGG TNIGTATTT CATTGGAGCC CCAGAGGGGA	300
40	GAGGCACATT TAAATATCA GAATTAGATT AGCTTGTAGT TTGTACAATT GGGAACATAA	360
	TAGATTTCA TAAATTATGT GTGCCTTGTG GGAAGTGTCA ACTGTCTTTA TGTCTGCTTG	420
45	TAAAAGTTTC AAAATATGTT TTCCCTCAAA AAGGCAACGT TACTTCATTT GCTTGAATAT	480
	TATGATAGGA ATGCTTACTG ATATTACTTG ATAGTCATAT ATAGCCTAGG AAATTTAAC	540
	TATATATAAC TATAGCAGTA TTAATAATGA TAGTTGTACT TCTTTAAAAC ATTAAATTG	600
50	AGGAAACTTT AATGCTGTCT CGTGTACATT GCTTTACTAC AGTGAGGGGG AATATCCTTT	660
	AGATTGAGCC TCAATTACT GGTTAGTAGT ATGTGAACTC TGGTATAAAA ACGTAAACTA	720
55	GACAGTAGAG CGGATGAATT AAAATTGTAA ATTGCTACAT TGGCATTTC TACCTCCTTT	780
	TCTGTCAGAG TATTACTTTT TCCAGCATTT ATTCTTATTT GTGAGTAAAG AGGAAATGGG	840
	AACCTGAGGT TAAAATTGAC ATTTTGTTT CATTGAGAAT TTAAGCAGTA GGTACAGGAG	900
60	AAGTGACTTG TCACATTAAT TIGGTGCCCTA AATCTGTAAC TACAAGTTGT GATCGACATG	960

	TACAAAATGT CTAAGAAAGG TCATATGCTG AATATTTAC TTTTCCTGTA TAGTCTGCAT	1020
5	GATTGTTTC ATAAACCCAG CTTATTTCTC CAAAAAAGCA AAATGGCCT GTAATTTTA	1080
	AAGTAAAATA AACGTGCCAT TTTGTCTGCA ATCTATAATT TCAGGAAGTT ATTGRAAGTT	1140
	CTGACTCAGG GCTTTTAAC AGTTCAAGCA ATTGTCAAGTT ATATTTTGGAA AACTCCATCT	1200
10	GTGTAATTCT CCAGTGCCCTT GAAAGAATTAA TTAACCTGGC AACACTATTA AAACTTTATA	1260
	AAAGATGGTC TTTAGTGCAC GTGTATCATT ATATACACGT TTTAAAGTCA TATTGCTTAG	1320
15	CTTGTAAATA ATGATTCTGC ATGTGTGCTG GGTTTGGTA ATTCTTTAAA GGAAGTTTC	1380
	TAGATTTGCA CTTGATGTTT GTTTTTAAA AACTGATTAT TTATGGCCGT GACACTGTTA	1440
	CCAGAAAAGT AATTCTAATT AAGTTATTAT GCAAAGTCAT CTATAAGTAG CATCTGGAA	1500
20	GAGGAGATSG AGGCCACAGT TTGCTATTTT AGTATGAAAG GAGGATCTGT TTGGGAAACA	1560
	TAGATTGTCT TCCCCTCAAA TGAGGGAAA AAAAAAGACC CTTTGTCAA ATGGATTCTG	1620
25	TTGTAAAAAA TTATTTTAA AGGAAATCAC AAATTGTATG TCATTCTTAA TGCTAGTCCT	1680
	ATAGAATAAA TCCATAAAAT TGTTTTATG TTCAGTATGT TTATGTCAATTCTAATGCAG	1740
	CAAATTCAAT GATAGCAGTT CAATTGACTC ATAGCAGTGT TTGTATTTT TTCTAATTCT	1800
30	TTAGCTTTCA ATATTGGATT AAAGTCTTGT TTGTGAATAT AGTTCCGTA TGCCAAATGA	1860
	TTTCTTGCTT ATTAGCTTTT GTAAAGAAT GCTTAGTAAG AGCTAAGCTT TTAAAAGTAA	1920
35	TGCAAACATT TATCGTTAAT AAAACCTATG GTGTAATATC ATATAATGCT TTTCTTGAT	1980
	CTTTGGAGAA TTATTCTTT ATAGTAGTAT ACATGAATT TGATTTTAA AGCATTTAAA	2040
	AACAAATCTC AATACATTAA AAAACCTGTT ATTGTAAAAA RGAAATTAC CATGCCTTAA	2100
40	AGAAACAAGG ATGTACATCT TCAATTCAAGC ATRAGTGTCC ACATCTAGAA GGCTCTCATT	2160
	GCAGTTGTTT ACAGTTAAGG TACCTCTATC TAAAGGCCA AAGAAGCATT TCATAYTTA	2220
45	ACACCTCACA TTCTTCAGG ATTAAGACAT ATGAAAATAG TCTGAATAGG ATAAATTG	2280
	ATAGGAAGTA ACTTAACCAAG TCTGGGAAGA TTCAGGCTTT TTCTATKAAA AAGCTTATTC	2340
	CTCTTCACAA CTCNGGTGGT AGGNTTTCAT TTTCAAGAG GGTAGATATT TTAAAGCCA	2399
50		

(2) INFORMATION FOR SEQ ID NO: 152:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 802 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

	CGTGCCTGTA GTAAGCTCAT CCCTGCCCTT GAGATGGTGA TGCCTGCCAA GGACAATGTT	60
5	TACCACTGG ACTGCTTTC ATGTCACTT TGTAATCAGA GATTNTGTGT TGGAGACAAA	120
	TTTTCCCAA AGAATAACWT GAYCCCTTGC CARACGGACT ACGAGGAAGG TTTAATGAAA	180
10	GAAGGTTATG CACCCCMGGT TCGCTGATCT ATCAACATCA CCCCATTAAG AATACAAAGC	240
	ACTACATTCT TTTATCTTTT TTGCTCCACA TGTACATAAG AATIGACACA GGAACCTACT	300
	GAATAGCGTA GATATAGGAA GCCAGGATGG TTATATGGAA TAAAAGGCGG ACTGCATCTG	360
15	TATGTAGTGA AATTGCCCA GTTCAGAGTT GAATGTTAT TATTAAGAA AAAAGTAATG	420
	TACATATGGC TGGATTTTT TGCTTGCTAT TCGTTTTGT GTCACTTGGC ATGAGATGTT	480
20	TATTTGGAC TATTGTATAT AATGTATTGT AATATTTGAA GCACAAATGT AATACAGTT	540
	TATTGIGITA CCATTTGTGT TCCATTTGCT YCTTTGTATT GTGCATTTA GTACAATCAG	600
	TGTTTAACT TACTGTATAT TTATGTTTC TGTATTTACC AGCTATTTA AATGAGCTGT	660
25	AACTTTCTAG TAAAGAATTG AAAAGCAAAT CCTCACTAAA GGATACACAG GATAGGATAA	720
	AGCCAAGTCN CATCAACATT AAAAAAACT AAAANANAAA ACACAAAAAA AAAAAANCCC	780
30	GGGGGGGGCC CGGAACCCAT TC	802

(2) INFORMATION FOR SEQ ID NO: 153:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH:-461 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

	CTAGGACCAC CGAGCAGCTT GGCTAAAAGT AAGGGTGTG TGCTGATGGC CCTGTGCGCA	60
45	CTGACCCCGCG CTCTGCNCTC TCTGAACCTG GCGCCCCCGA CCCCTGCCGC CCCCTGCCCG	120
	AGTCTGTTCC CGCGCGGCCA GATGATGAAC AATGGCCTCC TCCAACAGCC CTCTGCCCTG	180
50	ATGTTGCTCC CCTGCCGCC AGTTCTTACT TCTGTGGCCC TTAATGCCAA CTTTGTGTCC	240
	TGGAAGAGTC GTACCAAGTA CACCATTACA CCAGTGAAGA TGAGGAAGTC TGGGGGCCGA	300
	GACCACACAG GTGGGAACAA GGACAGGGGG ATTAAAGCAG TCAAAAGGAA AAACATGTTA	360
55	AGACCCTAGA CTTGTATATT GACACACTTG TACCTTGAA GGCAGAGGAA TGTAATTAAA	420
	AAGCACTTAT TTGGCWAAA AAAAAAAAAA AAAAAAAAAA C	461

(2) INFORMATION FOR SEQ ID NO: 154:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2388 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

15	GCCCCACCGGT CCGAAAGCGG AGAACCGCTGG TGGGCCTGTT GTGGAGTACG CTTTGGACTG AGAACCATCG AGGCTATAGG ACGCAGCTGT TGCCATGACG GCCCAGGGGG GCTGGTGGCT	60 120
20	AACCGAGGCC GGCGCTCAA GTGGGCCATT GAGCTAACCG GGCCTGGAGG AGGCAGCAGG GGTCGAAGTG ACCGGGGCAG TGCCAGGGGA GACTCGCTCT ACCCAGTCGG TTACTTGGAC AACCAAGTGC CTGATACCAG CGTGCAAGAG ACAGACCGGA TCCCTGGTGA GAAGCGCTGC	180 240 300
25	TGGGACATCG CCTTGGGTCC CCTCAAACAG ATTCCCATGA ATCTCTTCAT CATGTACATG GCAGGCAATA CTATCTCCAT CTTCCTACT ATGATGGTGT GTATGATGGC CTGGCGACCC ATTCAAGGCAC TTATGCCAT TTCAGCCACT TTCAAGAIGT TAGAAAGITC AAGCCAGAAG	360 420 480
30	TTTCTTCAGG GTTGGTCTA TCTCATTGGG AACCTGATGG GTTGGCATT GGCTGTTAC AAGTGCCAGT CCATGGACT GTTACCTACA CATGCATCGG ATTGGTTAGC CTTCTTGAG	540 600
35	CCCCCTGAGA GAATGGAGTT CAGTGGTGGA GGACTGCTTT TGTGAACATG AGAAAGCAGC GCCTGGTCCC TATGTATTTG GGTCTTATTT ACATCTCTCT TTAAGCCCAG TGGCTCCTCA GCATACTCTT AAACTAATCA CTTATGTTAA AAAGAACCAA AAGACTCTTT TCTCCATGGT	660 720 780
40	GGGGTGACAG GTCCTAGAAG GACAATGTGC ATATTACGAC AAACACAAAG AAACTATACC ATAACCAAG GCTGAAAATA ATGTAGAAAA CTTTATTTT GTTCCAGTA CAGAGCAAAA	840 900
45	CAACAACAAA AAAACATAAC TATGAAACA AGAGAATAAC TGCTGCTAAA TCAAGAACTG TTGCAGCATC TCCTTCAT AAAATTAAATG GTTGAGAACAA ATGCATAAAA AAAGTTGCAC AAGTTCCCTTA TTTTCCCTAA TATTTCACTT CTATTTAATA CAAGCTGGGA CATAAAAATT	960 1020 1080
50	CTGTTGGGA TACCTGGGG AAGATGTGAG AAACTAATGC TGAATTCAAGC TTATACATGA TGAAAAGAAA AACCAAGACAA AAGGAGCACAA TAAATATGCA TACAGTGTAA CTGTTATTAT	1140 1200
55	TTAATACCC ACGATAAGGG ATTTTGTTA GCATGTTAG GGGGAACGAG GATTGGTGGG ATCCTTGGGG CCACAGGAAT CTGAGGCAAC GGAAGATATA TAGAGTGATC GTCCCCCTGC CGAAGGAACC TGGCAYCTGT CAAGCAGATG CTGCAGTTCA AACTTCAGCT TTTAAGATAG	1260 1320 1380
60	ATAGCTATTG AAGGCAGAGG GTCAAGCAGGA GGATGTGTAT TTCTAATCTA CCCTGGTAAA	1440

	GTCATAGGTA AGACTCAAAA GCGGGATCTT ATTCAAAAGG CAGGTATTTC CTTTGTTTTC	1500
	TGTCTTGAAA TAGCCCCCTTC CCCTAAGGTG CATTCTCTCA AGTTTCAGT ATTGCTTTAT	1560
5	TTGCAGTGAT TAAAAGAGAT GAGAGACTTT GGAGACAGAC AACGTAAGCA ACACATACAC	1620
	ACATGAAATA CTCTAGACAG AGATGAATAT AAATCTGCC TAATAACCAG TTTTCCATGT	1680
10	AACAGTGATT TTGIGITTCG GGCTGAAGCA GTGGTTATAT TAAAAGCCAC TAATTCCCTT	1740
	ATCCCTTTAA AAGATTTTA CAATTCTCCA ACCACAAACA GCACTCTAA AACTAACTTT	1800
	ACTTTCTGCC CATAATTGT TCTACATGGA AAAAAAAAAT ATTACTTGG CCAGGGGTGT	1860
15	GTGTAAATGT GCCAGAATTC CTAGGCAGGC TGACCTTAC AGTATGGCC TTTAAGATAC	1920
	TGGATCCTGG TTGGGCAACA AGTGTACGC CTGAAGTTTC TGAAAACAAA TTAGAAGACT	1980
20	GTTGGCTTGG CTAATCTCGT AGTCAGGGC CAAGTTCTG TAGTCAGAAT GAAGAATAAA	2040
	ATTGAAAGAA AAAGGGGAA ATGCTTATAC TTGGCATTTAA GTGAAATGCC TCAAGTCTTA	2100
	ACTATGGCTT TGTAGATGAG GCAAAGATT TCTTAGTGGT AAAATTTCTT CAACAGGTCA	2160
25	ATGCCAATCT GTATGCCATT TTAGTAAAGT AGGTAAAGGAG AGTAGCCGCT CAGTAACTTT	2220
	GGCACTAAAG AAAGAGTGTG GCTCTAGAAC TTCCAATCCC ATTGCTAGAT GTGCCCTTTA	2280
	AAAGATGGTC CAGTGCTTTC AGGGAAGGAT GTTTAGCCAG TTTTCCTAGT ATTTGTTCC	2340
30	TAAGATTTT TGACCTGTGC TTAATAAGAC GGACCGGTGG GTCGACCC	2388

35

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 642 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

45	AAAACAGACC ATTAAAAAC TCAGACAAGA TTATATTAA TATATTAATT ACTAAAAAGG	60
	CACAAGATTA CACTGAACAT ATTAGCTACT AAAAAGGCAC TGCTAAGACA TTCAAGCAA	120
50	TAGCTATTAC ACACACTGC AGATTTACA GGTTCTAAT TCTAACATAT GTTGAAAAA	180
	TCCGTGAGTA TTCCAAAATA TATTTAATAA TGGAATATCT GCATTAATAT ACCATCCATG	240
	TGTTTTTACC ATTTGCCCTTA ATATTGAATA TACTGTTAC CTCACACTAA AAAGAAAACC	300
55	AGAAGCCTTA TTTGTGATTT TGGGAGTGGG AGCTTCCATT TTTGTGTCAA AAATGAATCC	360
	TGATTCTTAT GGAAATCTCT GTTATTAAGA TATTTCAAGA TGAGACAACA CTGAAGATCA	420
60	AATTGTGTTT AGTATCACTA TCTTCTCTCC TCGTTCTCT CTTACTCCCTC ATCCCTCCAG	480

AATCTACCAAG TTTATGGTAG AAAGATGGGA ACCTTATTG AATGTGTTT TTTTTTCCA	540
TGATGTCAA TTTGTGTG GGAAAGGATT TGGATAAAAT TTTGTTAA ATTTGGTAG	600
5 ATTTTATCT ATACAAATT AAATAAAATT ATGTTTGTA AG	642

10

(2) INFORMATION FOR SEQ ID NO: 156:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1251 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

20 GCGGCTGCC CTCCACGGAG TTGCTGATCA TCTGGGCTGT GATCCACAAA CCCGGTTCTT	60
TGTCCCTCCT AATATCAAAC AGTGGATTGC CTTGCTGCAG AGGGAAACT GCACGTTAA	120
25 AGAGAAAATA TCACGGCCG CTTCCACAA TGCAGTTGCT GTAGTCATCT ACAATAATAA	180
ATCCAAAGAG GAGCCAGTTA CCATGACTCA TCCAGGCACT GAGCATATTA TTGCTGTCA	240
30 GATAACAGAA TTGAGGGTA AGGATATTTT GAGTTATCTG GAGAAAAACA TCTCTGTACA	300
AATGACAATA GCTGTTGGAA CTCGAATGCC ACCGAAGAAC TTCAGCCGTG GCTCTCTAGT	360
CTTCGTGTCA ATATCCTTTA TTGTTTGAT GATTATTCT TCAGCATGGC TCATATTCTA	420
35 CTTCAATTCAAG AAGATCAGGT ACACAAATGC ACGCGACAGG AACCAAGCGTC GTCTCGGAGA	480
TGCAGCCAAG AAAGCCATCA GTAAATTGAC AACCAGGACA GTAAAGAAGG GTGACAAGGA	540
AACTGACCCA GACTTGTGATC ATTGIGCACT CTGCATAGAG AGCTATAAGC AGAATGATGT	600
40 CGTCGGAATT CTCCCCGTCA AGCATGTTTT CCACAAATCC TGCGTGGATC CCTGGCTTAG	660
TGAACATTGT ACCTGTCTTA TGTGCAAAC TAATATATG AAGGCCCTGG GAATTGTGCC	720
45 GAATTGCCA TGTACTGATA ACGTAGCATT CGATATGGAA AGGCTCACCA GAACCCAAGC	780
TGTTAACCGA AGATCAGCCC TCGGCGACCT CGCCGGCGAC AACTCCCTTG GCCTTGAGCC	840
50 ACTTCGAACT TCGGGGATCT CACCTCTTCC TCAGGATGGG GAGCTCACTC CGAGAACAGG	900
AGAAATCAAC ATTGCAGTAA CAAAAGAATG GTTTATTATT GCCAGTTTG GCCTCCTCAG	960
TGCCCTCACA CTCTGCTACA TGATCATCAG AGCCACAGCT AGCTGAATG CTAATGAGGT	1020
55 AGAATGGTT TGAAGAAGAA AAAACCTGCT TTCTGACTGA TTTTGCTTG AAGGAAAAAA	1080
GAACCTATTT TTGIGCATCA TTTACCAATC ATGCCACACA AGCATTATT TTTAGTACAT	1140
60 TTTATTTT CATAAAATTG CTAATGCCAA AGCTTGTAT TAAAAGAAAT AAATAATAAA	1200

ATAAAAAAA AAAAACCCG GGGGGGCC CGTCCCCAAT TGGCCCTATG G 1251

5

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2127 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

15	CGGGCGGGAG AGGGAAGCTG CAGCGAGAGG CGCGGATCTC AGCGCGGAG CAGTGCTTCT	60
	GGGGCAGGCC CCTGAGGGAG GGAGCTGTCA GCCAGGGAAA ACCGAGAACCA CCATCACCAT	120
20	GACAACCAGT CACCAGCCTC AGGACAGATA CAAAGCTGTC TGGCTTATCT TCTTCATGCT	180
	GGGTCTGGGA ACGCTGCTCC CGTGAATT TTTCATGACG GCCACTCAGT ATTTCACAAA	240
25	CGGCCTGGAC ATGTCCCAGA ATGTGTCTT GGTCACTGCT GAACTGAGCA AGGACGCCA	300
	GGCGTCAGCG CNCCCTGCAG CACCCCTGCC TGAGCGGAAC TCTCTCAGTG CCATCTTCAA	360
	CAATGTCATG ACCCTATGTG CCATGCTGCC CCTGCTGTTA TTCACCTTACCC TCAACTCCTT	420
30	CCTGCATCAG AGGATCCCCC AGTCCGTACG GATCCTGGC AGCCTGGTGG CCATCCTGCT	480
	GGTGTTCCTG ATCACTGCCA TCCGGTGAA GGTGCAGCTG GATGCTCTGC CCTTCCTTGT	540
35	CATCACCATG ATCAAGATCG TGTCATTAA TTCATTTGGT GCCATCTGCA AGGGCAGCCT	600
	GTGGTGTCTG GCTGGCCCTTC TGCCCTGCCAG CTRACACGGC CCCCATCATG AGTGGCCAGG	660
	GCCTAGCAGG CTTCTTGCC TCCGTGGCCA TGATCTGCGC TATTGCCAGT GGCTCGGAGC	720
40	TATCAGAAAG TGCCCTGGC TACTTTATCA CAGCCTGTGC TGTKATCATT TTGACCATCA	780
	TCTGTTACCT GGGCCTGCC CGCCTGGAAT TCTACCGCTA CTACCAAGCAG CTCAAGCTTG	840
45	AAGGACCCGG GGAGCAGGAG ACCAAGTTGG ACCTCATTAG CAAAGGAGAG GAGCCAAGAG	900
	CAGGCAAAGA GGAATCTGGA GTTTCAGTCT CCAACTCTCA GCCCACCAAT GAAAGCCACT	960
	CTATCAAAGC CATCCTGAAA AATATCTCAG TCCGGCTTT CTCGTCTGC TTCATCTTCA	1020
50	CTATCACCAT TGGGATGTTT CCAGCCGTGA CTGTTGAGGT CAAGTCCAGC ATCGCAGGCC	1080
	GCAGCACCTG GGAACGTTAC TTCAATTCTG TGTCCCTGTT CTTGACTTTA AATATCTTG	1140
55	ACTGGTTGGG CCGGAGCCTC ACAGCTGTAT TCATGTGGCC TGGGAAGGAC AGCCGCTGGC	1200
	TGCCAAGCTG GNTGCTGGCC CGGCTGGTGT TTGTGCCACT GCTGCTGCTG TGCAACATTA	1260
	AGCCCCGCGG CTACCTGACT GTGGTCTTCG AGCACCGATGC CTGGTTCATC TTCTTCATGG	1320
60	CTGCCCTTGC CTTCTCCAAC GGCTACCTCG CCAGCCTCTG CATGTGCTTC GGGCCCAAGA	1380

	AAGTGAAGCC AGCTGAGGCA GAGACCGCAG AGCCATCATG GCCTTCCTCC TGTGTCTGGG	1440
5	TCTGGCACTG GGGGCTGTTT TCTCCCTCCT GTTCCGGCA ATTGTGTGAC AAAGGATGGA	1500
	CAGAAGGACT GCCTGCCTCC CTCCCCTGTCT GCCTCCTGCC CCTTCCTTCT GCCAGGGGTG	1560
	ATCCTGAGTG GTCTGGCGGT TTTTCTTCT AACTGACTTC TGCTTCCAC GGCGTGTGCT	1620
10	GGGCCCGGAT CTCCAGGCC C TGGGGAGGGA GCCTCTGGAC GGACAGTGGG GACATTGTGG	1680
	GTTTGGGGCT CAGAGTCGAG GGACGGGGTG TAGCCTCGGC ATTTGCTTGA GTTTCTCCAC	1740
15	TCTTGGCTCT GACTGATCCC TGCTTGTGCA GCCCAGTGGA GGCTCTTGGG CTTGGAGAAC	1800
	ACGTGTGTCT CTGTGTATGT GTCTGTGTGT CTGCGTCCGT GTCTGTAGA CTGTCTGCCT	1860
	GTCCTGGGGT GGCTAGGAGC TCGGTCTGAC CGTTGTATGG TTTGACCTGA TATACTCCAT	1920
20	TCTCCCCCTGC GCCTCCCTCCT CTGTGTCTC TCCATGTCCC CCTCCCAACT CCCCATGCC	1980
	AGTTCTTACC CATCATGCAC CCTGTACAGT TGCCACGTTA CTGCCTTTTT TAAAAATATA	2040
	TTTGACAGAA ACCAGGTGCC TTCAAGAGGT CTCTGATTTA AATAAACCTT TCTGTGTTTT	2100
25	TTCTCCATGG AAAAAAAAAA AAAAAAAA	2127

30

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1625 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

40	CAAAAGATCT ATAATCAGGA CATTGTTTAT GTAAGTTGGA CAANAAAAAT TCTTCCCTT	60
	TATGTCCACC CTTCCATGA TTGCAAGACA AAATTTCCT CCTTTACCTC ATCCCTATAA	120
45	CATGGGAGGC TGAGAAAAAT GAGGGGAGAT GGAACCAGAT ACAAGGAGAT CCAATAAGAG	180
	AAGCTTATTT AAATATTGTG AAATAAAGGA AGAMCAAAG CATTTCCTTA AGTGGGAAT	240
50	CCTTTTGAAC AGTTATTATT TATCCATATT ATTAAYAAC A TCTTTCTGA CAAATCCAT	300
	CAGATGAAGT GTAAATGGAT AATCTTTAA TGGATCTAA CCTAGAAAGT TTCACTTACT	360
	GTTCATGTCC GTGTTCCAGA ATTGTGAAAT GGTGTGTGGT TTTGCTTCC AAGTTCTTCT	420
55	CTGCCTCCCTC TTAATTCTCT AATTCCATGT CTTACAGAAAG AATGAGAAAT TTCTTTCTTA	480
	CTTGAGTATC ATGCTCTAAA AAACCTGGCT TCAGTCACAG AAACGCTGGC TCTCTGTGC	540
60	TTATATTGAA GCCAACTGCC TTTAATTCTT GGGCCCTCTT ATATTTTAA GGTGCAAAAT	600

	TGAAAGTCTC AGTCACCAGA CACAGGTTCT ATACAATTAA TGATGACCTG GAGAAGTAAT	660
	ATGTAGCTAA TTTTCAAAAA GCATTGAATA TACTTTCCGG AAAGAAAACA GAAATTAAAT	720
5	ATTGCCACAT CTTGCCAGAA TCCCACATCTGA CACCTTAACCT TTGTCAGGTT TCCTACAAC	780
	TGCTAACCAA GTTTTATACA TTCTAAATCT CCCCAGTTTC TTTGGGCTG GAAGATGCAA	840
10	CTTCCATTAA ATAGAAACTT TGAAATCTTG GGGTAAGGG ACGTAGGGGG GACTAGGGAG	900
	AAGGATAAGA AATAGAATTAA TTGAAAAGCC CCCACCAGGG ACCCTCCCTGG CCAGAAATATG	960
	CAGAGTAATT CCTGCTGGCT TCACCTTTGA AAGTCCCTCG AACTATGCA GATGAAACTG	1020
15	AGTCTGTTTT TGATATTGTC AGATGTAATTC TACCTTGAA GTCCCNACAC CTAAACTGGA	1080
	ATTCTTGTAT TTACATCTCC TCCACTGTCC CCCACACCAC CCCTCAATTTC CTGCTGCC	1140
20	TGCTAACCAA AAGCATTCTTCTCTTGTAT CATCAGGTT ACATTAAM CAGTACTTA	1200
	CAAACGTACT TGAAGCACAG ATACCTTTAC GAATGTGATA AAATATTTTC TTAAGAAAAG	1260
	GAAAGAGGAT GTGGGTCAAA TAAAACACCG CATGGATGTT GATTGGTGA TACTGGTGT	1320
25	AGAAAAGGG A GTCAGGAAT TTTTATTACT GTATTGTAA ATGAGTTGAA AGGAATTGT	1380
	AAATGCCACT GGTACATTAA TAAGGTGACA CATTGCTCC TTATAAAGTT ATTAAAATT	1440
30	ACAGGGTAAG CTTAAATGAC GTTGCAGT AGTTTACTT TATATAATCA ATATTGATAT	1500
	TGTTGCTGAA CTATGTAAC TTATGATGCA TTTTCAGTC CCTTTTCAGA GCAAATGCTT	1560
	TTGCAATGGT AGTAATGTT AGTTAAATT GACTTAATAA ATTTTACCT GAGCAAAAAA	1620
35	AAAAA	1625

40 (2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1687 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

50	CGGGGTCAACC AGTTATTAGA GGAAGTAACA CAAGGGATA TGAGTGCAGC AGACACATT	60
	CTGTCGATC TGCCAAGGG TGATATCTAT GTGTCAGATG TTGAGGACGA CGGTGATGAC	120
55	ACATCTCTGG ATAGTGACCT GGATCCAGAG GAGCTGGCAG GAGTCAGGGG ACATCAGGGT	180
	CTAAGGGACC AAAAGCGTAT GCGACTTACT GAAGTGCAAG ATGATAAAGA GGAGGAGGAG	240
	GAGGAGAAC CACTGCTGGT ACCACTGGAG GAAAAGGCAG TACTGCAGGA AGAACAAAGCC	300
60	AACCTGTTGT TCTCAAAGGG CAGCTTGTGCT GGGNATCGAG GACGATGCCG ATGAAGGCC	360

	TCGAGATCAG TCAGGCCAG CTGTATTIG AGAACCGGYG GAAGGGACGG CAGCAGCAGC	420
5	AGAACAGCA GCTGCCACAG ACACCCCTT CCTGTTGAA GACTGAGATA ATGTCTCCCC	480
	TGTACCAAGA TGAAGCCCCT AAGGNAACAG AGGCTTCTTC GGGGACAGAA GCTGCCACTG	540
	GCCTTGAAGG GGAAGAAAAG GATGGCATCT CAGACAGTGA TAGCAGTACT ACCARTGAGG	600
10	AAGAAGAGAG CTGGGAACCC TCCGTGGTAA GAACCGAASC GTGGGCCTAA AGTCAGATGA	560
	TGACGGTTT GAGATAGTGC CTATTGAGGA CCCAGCGAAA CATCGGATAC TGGACCCCGA	720
15	AGGCCCTTGCT CTAGGTGCTG TTATTGCTC TTCCAAAAAG GCCAAGAGAG ACCTCATAGA	780
	TAACCTCTTC AACCCGTACA CATTAAATGA GGATGAGGGG GAGCTTCCGG AGTGGTTTGT	840
	GCAAGAGGAA AAGCAGCACC GGATACGACA GTTGCCTGTT GGTAAGAAGG AGGTGGAGCA	900
20	TTACCGGAAA CGCTGGCGGG AAATCAATGC ACGTCCCATC AAGAAGGTGG CTGAGGCTAA	960
	GGCTAGAAAAG AAAAGGAGGA TGCTGAAGAG GCTGGAGCAG ACCAGGAAGA AGGCAGAAC	1020
25	CGTGGTGAAC ACAGTGGACA TCTNCAGAAC GAGAGAAAGT GGCACAGCTG CGAAGTCTCT	1080
	ACAAGAAGGC TGGGCTTGGC AAGGAGAAC GCCATGTAC CTACGTTGTA GCCAAAAAAG	1140
	GTGTGGCCCG CAAAGTGGCGC CGGCCAGCTG GAGTCAGAGG TCATTCAAG GTGGTGGACT	1200
30	CAAGGATGAA GAAGGACCAA AGACCACAGC AACGTAAGGA ACAAAAGAAA AAACACAAAC	1260
	GGAAGTAAGC AGAGCTGCCA GGCTCCCAGG AGAGCATGGG GACTAGGAGG AAGGGTGTGG	1320
35	CATGGCTCAG TCTGGCCCCC TTGATTACCG GCCTAGCCCC TGTCACATC ACAGCTGTCT	1380
	GAAGAACAGT GAGGTGGAGT GCCTAGAACT CCCGTGGTGG TCCGTAGCAG AGAGGAGGAT	1440
	GTCCTCCTGC CTGCCTGAAG GTCTCCCCTG AAAACACTGC TGAACGTGT TGACACTCCT	1500
40	GACCCTTTTT TTAAACCGTT AAAGGAAAGT TCGGTGTGG AGCGATACTC AATGTAGTCA	1560
	GTCTACACCT GGACGTGTGG GCCACTTAAG CCCTCCCCAC CCCCATCCTA TTCCCTRAATA	1620
	AAACCAGGAT AATGGAARAA AAAAAAAAAA AAAAAAAAAG GGGGGGCCN TAAAGGGNCC	1680
45	CANNTTT	1687

50

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 1842 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

	GGATGACAGA TTGCGACANA GATTTGIGAC CCTTCCTGCT GAACTTCAGA GGGAGCTGAA	60
	ANCAGCGTAT GATCAAAGAC AAAGGCAGGG CGAGAACAGC ACTCACCCAGC AGTCAGCCAG	120
5	CGCATCTGTG CCCCAGAAT CCTTTACTTC ATCTAAAGGC AGCAGTGAAA GAAAAGAAAA	180
	GAAACAAGAA GAAAAAAACC ATTGGTCAC CAAAAAGGAT TCAGAGTCCT TTGAATAACA	240
10	AGCTGCTTAA CAGTCCTGCA AAAACTCTGC CAGGGCCTG TGGCAGTCCT CAGAAGTTAA	300
	TTGATGGTT TCTAAACAT GAAGGACCTC CTGCAGAGAA ACCCCTGGAA GAACTCTTG	360
	CTTCTACTTC AGGTGTGCCA GCCCTTCTA GTTGCAGTC TGACCCAGCT GGCTGTGTGA	420
15	GACCTCCAGC ACCCAATCTA GCTGGAGCTG TTGAATTCAA TGATGTGAAG ACCTTGCTCA	480
	GAGAATGGAT AACTACAATT TCAGATCCAA TCGAAGAAGA CATTCTCCAA GTTGTGAAAT	540
20	ACTGTACTGA TCTAATAGAA GAAAAGATT TGGAAAAGT GGATCTAGTT ATAAAATACA	600
	TGAAAAGGCT GATGCAGCAA TCGGTGGAAT CGGTTGGAA TATGGCATTG GACTTTATTG	660
	TTGACAATGT CCAGGTGGTT TTACAACAAA CTTATGGAAG CACATTAAGA GTTACATAAA	720
25	TATTACCAAGA GAGCCTGATG CTCCTGTATA GCTGTGCCAT AAGTGTGTGAGT GAGGTATTG	780
	CAAAGTGCAT GATAGTAATG CTCGGAGTTT TTATAATTAA AAATTTCTTT TAAAGCAAGT	840
30	GTTTGTACA TTTCCTTCA AAAAGTGCCA AATTGTGAG TATTGCATGT AAATAATTGT	900
	GTAAATTATT TTACTGTAGC ATAGATTCTA TTTACAAAAT GTTGTGTTAT AAAGTTTTAT	960
	GGATTTTAC AGTGAAGTGT TTACAGTTGT TTAATAAAGA ACTGTATGTA TATTTGGTAC	1020
35	RGGCTCCCTT TKGTGAAAYCC TTAAAAACTC AACTCTAGGA RGCAACTACT GTTTATTATA	1080
	CTAAARGGCT GAAAAMCTC CAGGCCAGAC TGCTAAGCTC TGAAATYCCT GAGAGGTCTC	1140
40	AGACCGGGAT TCTACTGTGTT CCAAGAAAGG GTAAAGCTTC TAAACCATCT TATTCTGTGTC	1200
	TCCAAGCATG AACACAGGAG CATGTYAAGA AAATCTTAC TACTTTCTYC CATGCGGAGA	1260
	AATCTACATA TTTTGAATTAA GAAACACCCCT CACACCCACT TGAAGATTAA TTTCTGGGA	1320
45	ACATTATGTC CGTAGATCA GAGGTGGTGT TGTCTTTTG CTTCTACTGG CCATTGAGAA	1380
	ACTTTGATGA TAAAAAAGAA CGGTATAGAT TTTTCAAACG TATATAAAAT ATTTTTATGT	1440
	TATATGTTAT GCCATAACTT TAAAATAAAA ATAGTTAAA ATTCTATGCT AGTGGATATT	1500
50	TGGAACCTTT TCCTCAAACA AACACCCCCAC ACTGACTTCA GCAAAACCTT AAAACTAGCT	1560
	ACAGATTACT ACTACGAATG AATCATYAAG TTTTGTGTCT GCAACAATTG AGAAGCACTA	1620
55	AGCCCAAATA TCAGGAAATG TGTGTATGAT GGAATTCTCT AGGACAAAAC AGATCAAGAT	1680
	TAAAACAGGA TCAAGGATTA ATGGTATAAA AATGGTCTAC TAAAACAGGA TCAAGGATTA	1740
60	AAACAGGATC AAGGATTAAT GGTATAAAA TCTCTACTGG TTACCGGGTG CCNGGGCCAT	1800

ACAGGGTAGT GGTGGATGGA TAGTTAGTT TGGNAAGGGT AA 1842

5

(2) INFORMATION FOR SEQ ID NO: 161:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 770 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

15	GGCACGAGCC CTATGCTGTT CTTGTGATAA TGAGTGAGTC TCACAAGATC TGGTGGTGT	60
	ATAGGCATCT GGCAATTCCC CTGCTGACGC TCATTCTCTA TCCTGCCACC CTGGGAAGAA	120
20	GTGTCTCTG TCATGATTGT AAGTTCTG AGGCCTCCCC AGCTATGTAG AACTGTGAGC	180
	CAATTAAACC TCTTTCTCT ATAAATTATC CAGTCCTATA TAATTCTICA TAGCAGTGT	240
25	AGAACAGATA ATACCGTAAA TTGGTATCAC AGAGAGTGGG GTGTIGCTAT AAACACATCT	300
	GAAAATGTTA AAGCAAATTG CGAACTGGGT AACAGGCAAA GGCTGGAACA GTTKGAAGAA	360
	CAGTTAAGAA GAAGACAGGA AAATATGAGA AATCTTGAAA CTTCTAGAG TCTTAAAGGT	420
30	CTCAGAAGAC ATGAAGATGT CGGAAGCTTT GGAACCTCCT AGAGACTTGT TTGAATGGCT	480
	TTGACCAAAA TGCTGATAGT GATATGGACA ATGAAGTCCA GGCTGAGCTT ATCCAGACAG	540
35	ACATAAGAAG CTCGCTGGGA ACTTGAGTAA AGATCACTCT TGCTAGGCAA AGAGACTGGT	600
	GGCCTTTTTT CCTCTGCCCT AGAGATCTGT GGAAATCTGA ACCTGAGAGA GATGATTTAG	660
	GGTATCTGGC AGAAGAAATA TCTAAGCGGC AAAACCTTCM AGAGGAAGCA GAGCATAAAC	720
40	GTTGAAAAA TTTGCAGCCT GACNATGGGA GACCAAAGTT AAACCCAATT	770

45 (2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 519 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

55	GAATTGGCA CGAGCTGAGA GGCACAGGAG CAACAGCCAG TGCCCCCTGC AGAGGACCAC	60
	TGGGGTCACA GACTTCARAC CTGATGACCT GGGCTCAGAT CCCAGCTCTG CACCTACCA	120
60	CCGTGTGACA AGGTGTCTC TCTGAGCCTC AGTCACACAC TGCCTTAACG GTTGGGCCTC	180

ATGGAGCTGT TIGTGAAGGT TAAATGGAA GACATAAAGC ACTTACCCCA GAGCCAAGGA	240
CATGCTGAAT AGGATAATGG TGGCCTCCCT TGGCGCTGTG CTGGGCCAGG TGTGCCGAGG	300
5 AAYTGGGCAG GGGTGACAGA TACCTCTTCT AACCTAGTTC CTTTCCAAGA ACCTAATTGG	360
TGTCTCTCCC TCCCCCAGGC AATTGGAAGG AGGAGGCTGG GCCCCAGCCC CAGAATACGG	420
10 GAGGTTTCTC ACCGTGGTAG GGAAATTGCT GGTTGGGGG TGTGGCAAC CACAGTGATC	480
GTCTCTCTGC AGGACGGATG AGGCTTTGCT GACAGAGGC	519

15

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 753 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

25 GGCACGAGCG GCACGAGCAG CCAGTTGCTG ACTGGCACAT GGCCCTCCAGC GTCCCCGGCTG	60
GTGGGCACAC TAGAGCCGGA GGGATCTTCT TAATTGGTAA ATTGGATCTT GAAGCTTCAC	120
30 TGTTTAAATC TTTTCAGTGG CTTCCCTTTG TACTTAGAAA AAATGCAAC TTCTTCTGCT	180
GGGACTCATC CGCTCACAGC CTTCCCCCTCC ACCCTCTCTC TGCCTCATGC TCTGCCCTTG	240
35 CCTGCCATGC CTCCGATACT CACCTTTTGT ACCCCAGCAC CCGTGCCCTC TGCCCCCTCGA	300
TCTTIGCTTG GCTGGTTGCT CCTCACTCAG TGTTCAGGAC AAATGCTCTT GGCCCTACCC	360
CATCTAGCCA GTCTAGCCCG GTCTTCCCTG TCTTCCCTGT TTCATTCAAG GCTCTTATTG	420
40 TTGTTWACT TGTGTGCTGT TGACTTTTAA CTCTCTCAGT CCCCACGTGA ATGCAAGCGA	480
TCTCCCAAGC TCCTAGAATT GTTCCCTGCCT CTTCACAGGC CCTTAACGCTG TGTGTGCTCG	540
45 TGGCGAATTG GGCACGAGGG TATGTGCACT TGCTGGTATG TATGTAGGTG TTTGCTAACAA	600
CATACTGCA CACGCAGAAT CCTTCCAGGG GACTGCACAG CCTCTAGTTC GCAGCCCCCA	660
CCCCCTCCCTT TGSCCCCTGCA CTCTCCCCCTC TCTGAGCTGC ATTGGCATGA AAGGGTGCAN	720
50 GGTTCCCTGAN CCCGCNAGCG NCACCTCCCTG GGA	753

55

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1400 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

5	GGCACAGTTT ATTAATACCT ATTATEGGAA AGTCACTTTC GTTGGCATTG AAAATTACAT	60
	CATCTTTAAA GCAGTATTG TCCCCAGNTG GACTCATCAC TAGCAAAGAC TAGGTTCAATT	120
10	GGAAGGCATA CGGTGAGAGA ATGGGAAGAT GRAGTGGAGG CGGGTTGTTA AAGTGCTGTC	180
	AGTGAGTGAT TTGTTCTACT TGAATANTGG TCCATGTTG GGGGCATATT GTGTTTCATA	240
	AGAAGTGAAA CGTATTTGCA AAGTAAAGCTA CAAATGACCC ATAAATCTGT TAACAACAGT	300
15	CCTTAATATG CAAAGATGAA AAACAGCT TACTGCTACC CAAAGGAAAC TGGTGCTTGG	360
	TGATGTGCAG ATGGGGCTGT TGGTTAAGCG AGCTATTACA GGTTTCTCT CTTAGGTTTC	420
20	ATAGGAGGTA GTTACTGAGA TGAGATTGTT TTATCTTTTT GAATACAGAT CTCTTGTCTT	480
	GAGTTAGTTC TGAGGATGGG AGTATAAAG GAGTTTTTG TTTTTTGTGT TGTTTGTTC	540
	TTTTGGCTCC TTAGTAATAC TCCTCTGACA TTTATTTCTA TTATTCTCA AAGAAAGGAA	600
25	ACCAACTGAA ATGTTGCTT TAACAAACAT TTTAATAAGT TCTCTGGTT TTTTTTGC	660
	CTTTAAAAAA AATTAGCATA TACCATACCA ATAAAAGAAC TAATGTTAAC TATTGTATGC	720
	TACAACCTAA GTGATTTTC TAAGAAGCA CAATGTCATT GRAAGTATTA TTGAAAAGGA	780
30	TCATAGTCAC ATTGAATTG TGAGGCCAA AGAAATTGAA CGGAGTGATA TTTTCATTTC	840
	ATGATATTCA CATATTTAGT AATTTTGTG TACAAGAATA CCAGGCAGAG TGTTTACCC	900
35	ATGGAAACAG GTTTCAGATT ACTTTGTTT TACTGTTAGA GTCTCAAGTT TAGAAATGCT	960
	AACACTTAAA TCAGTTTTT TCTCACTATA CTTGAAGATT GTTAATATT TGATATCTC	1020
	CTAGCTTGAT GGAATTAAA CATACTTCA GATCTGTGAC AGTGACAGCC AATAGGACTG	1080
40	ATAATATTAG CTTCAAACCA ATAATATCCA GGGTTAAAAT AAAAATCATA GTGAAAGTAC	1140
	GATTGTAAAA TTATGCTATA TTAACCTTTA AGTCIGTAAT AACTTGACAT CAAATGTTA	1200
45	TGTAATTACC ATAAATAATG GCTAGCGAGA ACATCTTTGG AAATTCTCAA ATTACCTTTC	1260
	TTACTACACT GTTTGCAGAA TGAATGTTAGA AATGATCCTG TTAGCTTCT GAATGTTCTG	1320
	TGGTGAATG TGTTTTGCT TAAATAAAGC TTTTGGTATT TGTTAAATW ACAAAAAAAA	1380
50	AAAAAAAAAA AAAAACTCGA	1400

55

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2153 base pairs
- (B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

5	CAGGCCTCAG GGCTCTGGT GGCTCTGCC CAGACAGTAT TTGCAAGTTCT TGTGCTATGG	60
	GTGGGAGTCT TCTTCCTCAA GTTTCGGCAG CTGIGCTGTG NCTGGATGGG CTGCTCCCTCC	120
10	CAGGGCTCAA GGGCTGTGGT CGGCTCAGGG TCTCATTTC CCAGGCCAAG TTCAAGGCAG	180
	CAGCCCTTG TGAGGGCGTC TTGGCCCTGG CCTGGAGGGA GAACTTTAAG CTTTTTGCT	240
15	CACAGGGACG TGGTATGGC CCTGGGTGCA GGTGCCACCA TTCTGCTAAT GAGAGCTTG	300
	TCTGATCAGT CCTGGGTCCA TCAGTTGTC CATGIGTCCG GCTGCCAGCC CGTCCCTTGG	360
	GATCCTTCCC CTGGGGTGA GCCTTGTCA TTAGTATATA CTCATTCCCT CATGCTTTCC	420
20	TCAGCAGAAC ACTTCCACTT CTGAGGTGAG CTTTGCCCC RTGCCCTTCC TCCACAGGTG	480
	TTGCCTTTT ATAAAGACCT GATAGCAGAA TAAATTGGTG TTTCCCTGTT GACCCAGCAC	540
25	CATTCTGTG GGCTAGAAT ATGCCCTCA ACCCTTAGAG TGGGGCAGTG AGGGCTTGAG	600
	GAGTGACCCCT TCCCTTCCTCA TGGTTTTAGT CATTTGGCT GCCAGCCCTT AATGGCACAG	660
	ATCTGCTGCT TCTAACAGAT GGCCAGGAGG TGACACOGAT TTCAGCCATT GCCAAGGTTA	720
30	GCACCCCTCTC CTTTGAGCCT AGGGCCACAC TGTTCATTGT CACTTGTAGC AAGTGCCTGT	780
	TTGGCTTTAA AGGTAAGCCT GCCAGCTGTG AGAAGCCTTG GTAAGTGATG GACTCATTTC	840
35	CTGGTCCCTTA AAGATGCAGC CTCTTAAGGG CTCCCTGATG GATGCCATCT CTCCCTAGCCC	900
	CCAGCCCTGG TGCCACTGGT GGGCAGGTTG CCATTCTTG GGGCTGGGAG GGACAGCTTG	960
	CCTGTTCTG GTCACAAATT ACAGTCTTCT CTCCTGTACC ATTCTGTGGC TTCAGCATGG	1020
40	GGGCAGTAGC CTTTCATTAG TGTAGATAGT CATTCCCTGG TAGGGTGGAG GGTAAGACAT	1080
	AGGGTCTGGA ACTGTTGGG ACCTTTGGG GATGTCCTGT GCCTCCAGA TTCCCTMGATT	1140
45	CTGGGAGGAG AGGCTGCCGC ATTCTGCTGC TCCACACAGC GAGCAAAGCT GCACCCACTT	1200
	ACATTCAAGTA TTTTCCCTGGC ACTACAAAGA GTGGGAAGGC CTGGGATTG CTGCTGCTCC	1260
	CTTAGAGCAG GGCCCCYYT TTCAGCACTT TGGACACCTG GAGACCCAGC CCTGTTATTT	1320
50	AATGGTAGTG GGCAAGTGTG TGTGCATACT GTCTGCCACT GCTTCTCCC TGCCCCATGC	1380
	CAGAGAGCCC TGTCCCTGCC AGGCCAGCC TTCTTAGCCC CAACTGGGA ACAAAAGTGCA	1440
55	ACATGGGATC ATGGGTTGGG GTGCTCAGGT GAGCCCTCTC TATACTGCTT CCCTGGGCCA	1500
	AGCTGACACC AGCCCCCTGAG GGTGGGGTGG GACGGGTGGT CCTTAAAAGA GGAAGGGGAC	1560
	CAGTGTAGCA ACTTGCCAGG GACCCACCC CTCCCTCTCT GGGCCGTGTC AGTGAGCATG	1620
60	GGGATTCCCCA TCAAGGGGCC TGGCACCTGT GCTAGTTACG TAGCCGCTGN TCAOGCGCTC	1680

	ACTCCTGACC ACATGCACGT TCCCTAGATG CAGACTGCCT TGAACTTTAA AGCTGTACAA	1740
5	TTTGGTTATG TTGTGCTGA CTAAAATAT ATTTTAATGA GGAAAAAATA ATGGAGAAC	1800
	CTGGGAAGGA CCTGGTCTT TTGCTCTCG GGGAACTGTA AGCCCTCGCG TTCTGGGAAT	1860
	CGCTCTCTGC TGCTCTTCC TGGAAGCTAA GCCTGTCCTC ACCGCCGAG GCCTGCGCCG	1920
10	GTGCTCCCGC CGCAGTTGCG TTTGCTTGG ACCTTGCGTG CGGGGGAGGG GGTGCTCGGT	1980
	CCGAGCCCGC TCCTTCTGT ACACCTAGCG CTGCCCCCCC CGCTGTGTC TGAGGTCGTG	2040
15	TATGTCAAAA ATAAAGCCGC TAGAAACGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	2100
	AAACTCGAGG GGGGGCCCGT ACCCAATTAA CCCNNTATGA TCTATAAAGC GTC	2153

20 (2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1251 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

30	GCCCACGCGT CGGCCACGC GTCCGGCGGT CGGGAGTATG GGGCGCTGAT GGCCATGGAG	60
	GGCTACTGGC GCTTCTGGC GCTGCTGGGG TCGGCACTGC TCGTCGGCTT CCTGTCGGTG	120
35	ATCTTCGCC CGTCTGGGT CCTCCACTAC CGAGAGGGC TTGGCTGGGA TGGGAGCGCA	180
	CTAGAGTTA ACTGGCACCC AGTGCTCATG GTCACCGCT TCGTCCTCAT CCAGGGCATC	240
40	GCCATCATCG TCTACAGACT GCCGTGGACC TGGAAATGCA GCAAGCTCCT GATGAAATCC	300
	ATCCATGCAG GGTTAAATGC AGTTGCTGCC ATTCTTGCAA TTATCTCTGT GGTGGCCGTG	360
	TTTGAGAAC ACAATGTTAA CAATATAGCC AATATGTACA GTCIGCACAG CTGGGTGG	420
45	CTGATAGCTG TCATATGCTA TTGTTACAG CTTCTTCAG GTTTTCAGT CTTCTGCTT	480
	CCATGGGCTC CGCTTCTCT CCGAGCATTT CTCATGCCA TACATGTTA TTCTGGAATT	540
50	GTCATCTTG GAACAGTGAT TGCAACAGCA CTTATGGGAT TGACAGAGAA ACTGATTIT	600
	TCCCTGAGAG ATCCTGCATA CAGTACATT CCGCCAGAAG GTGTTTCGT AAATACGCTT	660
	GGCCCTCTGA TCCGGTGTT CGGGGCCCTC ATTTTTGGA TAGTCACCAAG ACCGCAATGG	720
55	AAACGTCTTA AGGAGCCAAA TTCTACCAATT CTTCATCCAA ATGGAGGCAC TGAACAGGG	780
	GCAAGAGGTG CCATGCCAGC CTACTCTGGC AACAAACATGG ACAAAATCAGA TTCAGAGTTA	840
60	AACAGTGAAG TAGCAGCAAG GAAAAGAAC TTAGCTCTGG ATGAGGCTGG GCAGAGATCT	900

ACCATGTAAA ATGTTGAGA GATAGAGCCA TATAACGTCA CGTTTCAAAA CTAGCTCTAC	960
AGTTTGCTT CTCCTATTAG CCATATGATA ATTGGGCTAT GTAGTATCAA TATTTACTTT	1020
5 AATCACAAAG GATGGTTCT TGAAATAATT TGTATTGATT GAGGCCTATG AACTGACCTG	1080
AATTGGAAAG GATGTGATTA ATATAAATAA TAGCAGATAT AAATTGTGGT TATGTTACCT	1140
10 TTATCTTGTG GAGGACCACA ACATTAGCAC GGTGCCTTGT GCAKAATAGA TACTCAATAT	1200
GTGAATATGT GTCTACTAGT AGTTAATTGG ATAAAATGCC AGCATCCCTG A	1251

15

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 882 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

25 GACSMTCAG AACTATGGTC CCCCCGGACT GCAGGAATTG GGCACAGCGG CTGCGGGCGC	60
GAGGTGAGGG CGCGGAGGT CCCAGCAGGA TGCCCCGGCT CTGCAGGAAG CTGAAGTGAG	120
30 AGGCCCCGGAG AGGGCCCAGC CGGCCCCGGGG CAGGATGACC AAGGCCCCGGC TGTTCCGGCT	180
GTGGCTGGTG CTGGGGTCGG TGTTCATGAT CCTGCTGATC ATCGTGTACT GGGACAGCGC	240
35 AGGCGCCCGCG CACTTCTACT TGCACACGTC CTTCTCTAGG CGGCACACCGG GGCGCCCGCT	300
GCCCACGCCCGG GGGCGGACA GGGACAGGGG GCTCACGGCC GAYTCGGATG TCGACGAKTT	360
TCTGGACAAG TTTCCTCACTG CTGGCGTGAA GCAGAGTGAC YTTCCCAAGAA AGGAGACGGG	420
40 GCAGCCGCCT GGGCGGGGGG GCATGGAGGA GAGCGTGAGA RGCTACGACT GGTCCCCGGC	480
CGAMGCCCGG CGCACCCAGA CCAGGGCCGG CAGCARGCGG ANCGGAGGAR CGTGCTGCGG	540
45 GGCTTCTGCG CCAAYTCAG CCTGGCCCTTC CCCACCAAGG AGCGCGCATT CRACGACATC	600
CCCAACTCGG AGCTGAGCCA CCTGATCGTG GACGACCGGC ACGGGGCCAT CTACTGCTAC	660
GTGCCCAAGG TGGCCTGCAC CAACTGGAAG CGCGTRATGA TCGTGCTGAG CGGAAGCTGT	720
50 GCACCGCGTG CGCCTACCGC GACCCGYTGC GNTCCCCGGC GAGCACGTGC ACAACGCCAG	780
CGCGCACTGA CTTCAACAAT TCTGGCCGGG CTACGGGAAG TCTCCCCAC CTCATGAAGT	840
55 CAAGCTCAAG AATACACCAA TTCTTTCTGC GCGACCCCTTC TG	882

(2) INFORMATION FOR SEQ ID NO: 168:

60.

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1208 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

10	GGGAAACTCA AAAGGATGAT GGAATGGTTG ATGGAGCCAG AGCCTAGAAG TRAAGGGATA	60
	CAGAGTGAAG ATAGAGGTAT TTACGTATAT TTWAATATTA GCTTTGGAAT TACGTAGGGA	120
	TTCCTTAAGAA AAGATCATGA CAGGACAGCC ACATTTGGTA AAATGTCAGG GCAGCCAGTG	180
15	CATGGTCCTC CTGGGGCTCC TCAGTTGACG GGTTTAAATC ATTTCTGAT CCCCTGCC	240
	TGGTTTGAGG AATGCATACA GTACGTGAAA TGCCCTGTGGT ATGAGTTGCA ATGGGCAATC	300
20	AACCTGGGTA AATCCAAGAT TAATGATTAG TTCTAAAGAT CCAGTTGAAG TTCTAGAGTG	360
	GGAATTTTCC GTCAAGCARC TCAGCACAGC TTTATGCCCTG TTCCTCTAAT AACGATAGGT	420
	AACAAATAGC TGTGTTWCA CAGCTAGGAR GATAACAAA TCTAGAGTTC TTGARTCTCA	480
25	TTAATAAAAT AAKTATTATG AGTACCAACT GCATATTCA GGCACAGCAT TTGACTCTGT	540
	TAAATACTGA TYCCTTAKGA CMSCCACWTC AGAWAACMTT AATCTGTCTG ATCAATAAAC	600
30	AGCTTGACTT AGAGRGGTAA AATAGCTTGC CACAGGIWAC CCAATTAGTA GGTAACAGCG	660
	ACAGAATAAC AGTGCAGITA AAATCTTAGA CTGGAGACTA ATTGCATAAG TTTGAATTTC	720
	AGTTCTGCTA TGTAAATTG GGTGAGTACC TTAATTYACC TGAGTCTCGG TCTTTATATC	780
35	TGTAGAATGG AGCTAATGAT ATTACTTAAT TTGCTTTATG TGAGATTAAA TGTACTAATA	840
	TATGTAAATC ACTTACAACA GCATTTGACA TATTTGACAT ACTTAATATA TTTGCTACTA	900
40	ATACTATTAG CAACAGCATT CTGATTTTCC AAGTTGAAAT TCAGTGTGTTT CTTTTTACT	960
	TTGCCATAAT TTACAATGTT GTGCTCTGTA AACCATAAT TTCCCTGAGG TGTGTCAGG	1020
	TTAAAAAAA ATCACTATGG CCCCCARNMA CTTGGAAAAT AGAAATGAGA CCAGCTTCAT	1080
45	CTATATTCTT TACTGCAAAT AACTTAGAAT TGTAATAGGC TAATATGTAC TGGGACTTCC	1140
	AATTTGGGAA TATGACAAAA ATAATACTAT TTAGCTAAA CATATACAGA ACTTATTTC	1200
	CCTCTGAA	1208
50		

(2) INFORMATION FOR SEQ ID NO: 169:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1307 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 60 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5	GGCACGAGAG AAAAGAGGTT GAGAATGTTT TCTAGCAGGC AGAAATGTGCA TACATGTTT	60
	CATGARTGTC CTTTGGGTGC TGTTTCTTTT AAATCCCTTG TGCAACAGGGC TCTGGCCTTT	120
	ARTAAACTGT TTTTCTGTCT TACGTACATGC TGACTGGGTG CTAGGGCTG ATTACAAAGG	180
10	GGAAGAGTTG AACAGACATC AGGGGCCGAT GAAACCAAAG GACTAGGAGT CAGGAGAAC	240
	AGTCAGGGAT TAGGAGACAG CGGTTTGGTT TATTGTTATC CAGCTGGAGG ACTCCTAGGG	300
	GCAGCAGCAG GAGGAATACC AGGCCACGG AGGGGCAGGA GTCTCACAGT GGAGGGCAGA	360
15	CTCTAACAGA TGCCAGCTGA ACGCTCGCTG GCCCTGGATG TCATACGAGT TGGGGACCAG	420
	AAATCTGGCC TCAGAGAACCG CGTCCAGGGA GATTGAAAGC CATGGGTTAT CTTCTAGAGT	480
20	TGATACTGAT AATATATTTT AATTTTTATT GATGTTTAAT ACCTCTTGAA ACAGGAGGGT	540
	AAGATCAGAT GGGAAAGCCCY TCTGTTGAAG GATCTTGGGA ACCTTGGTGG TTTTTTTTTT	600
25	TTCGGTTTTT TTTTTTTGAT CGAGCTGTGG ACATCCTCT TAATTCGATT NTGAGGATT	660
	GTTTAACTAA AAAGTTCCCA AACACAGAAA GGGCCTCCCC ACCTGCTTTG GGGAGCTGTC	720
	TGTCTGGGA GTGCCAGGCA TCCSATGGGA CCCATCACTG CCAGTGTCTG TGCTCTCCAG	780
30	AGGTCAGCCC TGTGTCTGCC CTGGCTCTGT CTCCCTCTGTG ACAGGGCAGA GCATTTCTGG	840
	TCAGTTTCTC CATGGTGCCT CCCACCCCTT TGTAAAGTGG ATGGACATGA TTGAATTCA	900
35	TTGTCTCACC CTGATAGCCT GGGTGTGAT ATTCACTTTA CCCGCACTCA GACACAGGGC	960
	ACCTTGAAGC AGTTCTCGGT GTGTAGAGTC CACGTGACAG TCCCCACAGC CTCCCCAGAT	1020
	AGCTGTGTGC CTGTGCGCTA CTGCTGTGCC ATTTTCCCAA CTINGCGTT TCACTAAATG	1080
40	CAGCTGATCT CTCTCTCTGT GCACTCGTGA TCCATGTTGA ACAATACATG TAGTTCTTT	1140
	TTCCACGCAA TGTAAGAACAA TGATATACTG TACGTTGGAA AGCAATTACCC TTATTTATAT	1200
45	ACCTGAATGT TCCTACTACA CAANTAAACA TATATTAAT WCTAAAAAAA AAAAAAAAAA	1260
	CTGGAGGGGG GGCCCCGGTAC CCAAATCGCC GGATAGTGTGAT CGTAAAC	1307

50

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:

- 55 (A) LENGTH: 1624 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

	GGCACGAGGT CGCCGCCCGG GCGGCCTGGA ATTGTGGGAG TTGTGTCCTGC CACTCGGCTG	60
	CCGGACCGGA AGGTCCCTGA CTATGGCTCC CCAGAGCCTG CCTTCATCTA GGATGGCTCC	120
5	TCTGGGCATG CTGCTTGGGC TGCTGATGGC CGCCTGCCTC ACCTCTGCC TCAGTCATCA	180
	GAACCTGAAG GAGTTGCCCG TGACCAACCC AGAGAAGAGC ACCACCAAAG AAACRGAGAG	240
10	AAAAGAAACC AAAGCCGAGG AGGAGCTGGA TGCCGAAGTC CTGGAGGTGT TCCACCOGAC	300
	GCATGAGTGG CAGGCCCTTC ACCCAGGGCA GGCTGTCCT GCAGGATCCC ACGTACGGCT	360
	GAATCTTCAG ACTGGGGAAA GAGAGGCAAA ACTCCAATAT GAGGACAAGT TCCGAAATAA	420
15	TTTGAAAGGC AAAAGGCTGG ATATCAACAC CAACACCTAC ACATCTCAGG ATCTCAAGAG	480
	TCCACTGGCA AAATTCAAGG AGGGGGCAGA GATGGAGAGT TCAAAGGAAG ACAAGGCAAG	540
20	GCAGGCTGAG GTAAAGCGGC TCTCCGCCCG CATTGAGGAA CTGAAGAAAG ACTTTGATGA	600
	GCTGAATGTT GTCATTGAGA CTGACATGCA GATCATGGTA CGGCTGATCA ACAAGTCAA	660
	TAGTTCCAGC TCCAGTTGG AAGAGAAGAT TGCTGCGCTC TTTGATCTTG AATATTATGT	720
25	CCATCAGATG GACAATGCCGC AGGACCTGCT TTCTTTGGT GGTCTTCAAG TGGTGATCAA	780
	TGGGCTGAAC AGCACAGAGC CCCCTGTGAA GGAGTATGCT GCGTTTGTGC TGGGCCCTGC	840
30	CTTTTCCAGC AACCCCAAGG TCCAGGTGGA GGCCATCGAA GGGGGAGCCC TCCAGAAAGCT	900
	GCTGGTCATC CTGGCCACGG ACCAGCCGCT CACTGCAAAG AAGAAGGTCC TGTTTGCACT	960
	GTGCTCCCTG CTGCGCCACT TCCCCTATGC CCAGGGCAG TTCTGAAGC TCGGGGGCT	1020
35	GCAGGTCCTG AGGACCTGG TGCAGGAGAA GGGCACGGAG GTGCTGCCCG TGCCCGTGGT	1080
	CACACTGCTC TACGACCTGG TCACGGAGAA GATGTTGCC GAGGAGGAGG CTGAGCTGAC	1140
40	CCAGGAGATG TCCCCAGAGA AGCTGCAGCA GTATGCCAG GTACACCTCC TGCCAGGCCT	1200
	GTGGGAACAG GGCTGGTGCAG AGATCACGGC CCACCTCTG GCGCTGCCCG ACCATGATGC	1260
	CGGTGAGAAG GTGCTGCAGA CACTGGCGT CCTCCTGACC ACCTGCCGGG ACCGCTACCG	1320
45	TCAGGACCCC CAGCTOGCA GGACACTGGC CAGCCTGCAG GCTGAGTACC AGGTGCTGGC	1380
	CAGCCTGGAG CTGCAGGATG GTGAGGACGA GGGCTACTTC CAGGAGCTGC TGGCTCTGT	1440
50	CAACAGCTTG CTGAAGGAGC TGAGATGAGG CCCCCACACCA GGACTGGACT CGGATGCCGC	1500
	TAGTGAGGCT GAGGGGTGCC AGCGTGGGTG GGCTTCTCAG GCAGGAGGAC ATCTTGGCAG	1560
	TGCTGGCTTG GCCATTAAT GGAAACCTGA AGGCCAAAAA AAAAAAAAAA AAAAAAAAAA	1620
55	AAAA	1624

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2003 base pairs
 (B) TYPE: nucleic acid
 5 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

10	GGCACCGAGCC AGCTTGCAGG AGGAATCGGT GAGGTCCCTGT CCTGAGGCTG CTGTCCGGGG	60
	CCGGTGGCTG CCCTCAAGGT CCCTCCCTA GCTGCTGCCG TTGCCATTGC TTCTTGCCCTG	120
15	TTCTGGCATC AGGCACCTGG ATTGAGTTGC ACAGCTTTGC TTTATCCGGG CTTGTGTGCA	180
	GGGCCCGGCT GGGCTCCCCA TCTGCACATC CTGAGGACAG AAAAAGCTGG GTCTTGCTGT	240
	GCCCTCCCAG GCTTAGTGTGTT CCCCTCCCTCA AAGACTGACA GCCATCGTTC TGCAACGGGC	300
20	TTTCTGCATG TGACGCCAGC TAAGCATAGT AAGAAGTCCA GCCTAGGAAG GGAAGGATT	360
	TGGAGGTAGG TGGCTTTGGT GACACACTCA CTCTCTTCCTC AGCCTCCAGG ACACATATGCC	420
25	CTGTTTTAAG AGACATCTTA TTTTCTAAA GGTGAATTCT CAGATGATAG GTGAACCTGA	480
	GTTGCAGATA TACCAACTTC TGCTTGTATT TCCTAAATGA CAAAGATTAC CTAGCTAAGA	540
	AACTTCCTAG GGAACCTAGGG AACCTATGTG TTCCCTCAGT GTGGTTTCCT GAAGCCAGTG	600
30	ATATGGGGGT TAGGATAGGA AGAACTTTCT CGGTAATGAT AAGGAGAAC TCTTGTTC	660
	TCCCACCTGT GTTGTAAAGA TAAACTGACG ATATACAGC ACATTATGTA AACATACACA	720
35	CGCAATGAAA CGGAAGCTTG GCGGCCTGGG CGTGGCTTG CAAAATGCTT CCAAAGCCAC	780
	CTTAGCCTGT TCTATTCAAGC GGCAACCCCCA AAGCACCTGT TAAGACTCCT GACCCCCAAG	840
	TGGCATGCAG CCCCCATGCC CACCGGGACC TGGTCAGCAC AGATCTTGAT GACTTCCCTT	900
40	TCTAGGGCAG ACTGGGAGGG TATCCAGGAA TCGGCCCTG CCCCACGGGC GTTTCATGC	960
	TGTACAGTGA CCTAAAGTTG GTAAGATGTC ATAATGGACC AGTCCATGTG ATTTCACTAT	1020
45	ATACAACCTCC ACCAGACCCCC TCCAACCCAT ATAACACCCCC ACCCTGTTG CTTTCCGTGA	1080
	TGGTGTATATC ATATGTAACA TTTACTCCCTG TTTCTGCTGA TTGTTTTTTT AAATGTTTGG	1140
	TTTGTTTTTG ACATCAGCTG TAATCATTCC TGTGCTGTGT TTTTATTAC CCTTGGTAGG	1200
50	TATTAGACTT GCACCTTTT AAAAAAAGGT TTCTGCATCG TGGAAGCATT TGACCCAGAG	1260
	TGGAACGGGT GGCCTATGCA GGTGGATTCC TTCAGGTCTT TCCCTTGGTT CTTTGAGCAT	1320
55	CTTTGCTTTC ATTCGTCCTCC CGTCTTGGT TCTCCAGTTC AAATTATTGC AAAGTAAAGG	1380
	ATCTTTGAGT AGGTTGGTC TGAAAGGTGT GGCCTTATA TTTGATCCAC ACACGTTGGT	1440
	CTTTTAACCG TGCTGAGCAG AAAACAAAAC AGGTTAAGAA GAGCCGGGTG GCAGCTGACA	1500
60	GAGGAAGCCC CTCAAATACC TTACACATAA ATAGTGGCAA TATATATATA GTTTAAGAAG	1560

GCTCTCCATT TGGCATCGTT TAATTTATAT GTTATGTTCT AAGCACAGCT CTCTTCCTCCT	1620
5 ATTTTCATCC TGCAAGCAAC TCAAAATATT TAAAATAAAG TTTACATTGT AGTTATTTTC	1680
AAATCCTTGC TTGATAAGTA TTAAGAAATA TTGGACTTGC TGCGTAATT TAAAGCTCTG	1740
TTGATTTGT TTCCGTTGG ATTTTGAGG GAGGGGAGCA CTGTTTAT GCTGGAATAT	1800
10 GAAGTCIGAG ACCTTCCGGT GCTGGGAACA CACAAGAGTT GTGAAAGTT GACAACCAGA	1860
CTGCCATGT CTCTGATGCT TTGTATCATT CTTGAGCAAT CGCTCGGTCC GTGGACAATA	1920
AACAGTATTA TCAAAGAGAA AAAAAAAA AAAAAACTCG NGGGGGGCC CGGTACCCAA	1980
15 TTCCGCCTAT AGTGAGCCNA TTC	2003

20

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 786 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

30 GGCACAGCGG CACGAGAAGA CTTTGGTGTGTT TAAGAGATTA ATGTGTTAGC CAGAACAACT	60
CATTTCTCTA CCMGTGTGTA GTCCATTAT CTTTAAAGAT TTTCTATTGG AATAATTG	120
35 AAATTACTTT CTTAGTTTC TTCACTAAAA ACTAAGAAAA TGCTTGTGTT ATTATGAATT	180
GCTATTCTC TTGATTATTA TTCTTGGAGA AAGTCTATCA GACGTAATTG TTCTGATTG	240
40 CTTCTAGGCT AGAGGAAAAT GTGAAAGATG ACAAAATGAAA ATTCAAAAGG TTGTCAGTAG	300
TATGACTTCT TTTATCGTTT GTCAATTATCA CAAATATATC AACATAGGAC TTTTAAAAGA	360
TATTTGTAC ATATTGGCC TTAGTAGGAT TTTGCAATGAA TTTTTTTTTT CTTTATGCC	420
45 CAGAGAGAAA GAGCAAAGAA ATAACCAAGG GTGATGTAAT CGTATTGAAG GTTTACCAA	480
TAAGGACTGC TTTTATTATG AACTATAGTC TATATTCTAA GTAAATCAAT TTTCTATTA	540
50 TGTTGTTTTT GTCCCTGCAG GCAAGATCTC TGAACATTAT GCAGAGGGTT CTTTAAAAAA	600
AACAAAGTTG AATTTTTTTA TTCTTGGAA TATTTTTTTT CATTGATTTC TCCCAAGTAG	660
AGCAGATTCA AATCTCCCTT GTACCCATAG TCTTTTTTGT TTTGCTATTA GCTCAGTATT	720
55 CGTTTCTAC ATTTTCCCTT CCTAGAACCA GTCAATAAT GACAAAAAAA AAAAAAAA	780
ACTCGA	786

60

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1758 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

	GGGACGAGCC CTGCCAACCT CCTGCAGCCT CCTGCGCCCC GCCGAGCTGG CGGATGGAGC	60
15	TGCGCACGGG GAGCGTGGGC AGCCAGGCGG TGGCGCGGAG GATGGATGGG GACAGCCGAG	120
	ATGGCGGCGG CGGCAAGGAC GCCACCGGGT CGGAGGA CTA CGAGAACCTG CCGACTAGCG	180
	CCTCCGTGTC CACCCACATG ACAGCAGGAG CGATGGCCCG GATCCTGGAG CACTCGGTCA	240
20	TGTACCCGGT GGACTCGGTG AAGACACGAA TGCAGAGTTT GAGTCCAGAT CCCAAAGCCC	300
	AGTACACAAG TATCTACGGA GCCCTCAAGA AAATCATGCG GACCGAAGCT TCTGGAGGCC	360
25	CTTGCAGGGC GTCAACGTCA TGATCATGGG TGCAGGGCCR GCCCATGCCA TGTATTTGC	420
	CTGCTATGAA AACATGAAAA GGACTTTAAA TGACGTTTC CACCACCAAG GAAACAGCCA	480
	CCTAGCCAAC GGTATTTGA AAGCGTTGT CTGGAGTTAG AAAGTTCTCT TCTTCAACAC	540
30	GTCCTCCCC AGGGTGTTCC TCCCTGTGAC CCAGCCGCCT CGACTTCGGC CGCCTTGCTC	600
	ACGAATAAG AACTCAGAGT TGTGTGTGCA ATGCCACACCC AGACACACGC ACGCACACAC	660
35	ACGCGCGCGC ACACACATGC TTTTTCTGT TCCCTCCGC TTCTGAAGC CTGGGGAGAA	720
	ATCAGTGACA GAGGTGTTT GGTTTATTG TTATGTGGGT TTTCTTTGT ATTTTTTTTG	780
	TTTGTGTTGT TTTAACAT TCAAAAGCAA TTAATGATCA GACATAGGAG AAACCTGAA	840
40	TAGAAACAAA ACTTTGAAT GCTGGATTCA AAAAAAAA AAAGTTATCT GGACAGCTTC	900
	TTTGAGACTA TTTAAAAGT GGTACAACAG GTCTCTACAA CGCCAAGATC TAACTAAGCT	960
45	TTAAAAGGTC AAGAAGTTT ATGGCTGACA AAGGACTCGC GCAACGCAGA AGGCCTTCC	1020
	CACCTTAAGC TTCCGGGAT CTGGGAATT TACCCCCATT CTCTCTGTT TGTCTGAGTC	1080
	TCATCTCTCT GCAAGCAAGG GCTGAAATCA TTTTGTTGG TTGTTTGAG GGAGAGAGGC	1140
50	GGGGTGGGGG GGTCAAATC TGCCAGCAGC TCTTACGTA GGCAATGTTT ATGGGGAGG	1200
	GCTGAGCTTT TATTTCTCC TCTCCAGTGG GGTTGGCTTT TATTTGTTCT TGTTGGGTT	1260
55	TGGAATGGAA ATATGGATAG CAGCATAAAG TACTTTATT TTGACAAAAT TCATTTTTT	1320
	CAACAATGGA GACATAGATT TGACCCACAA TAACTCTCC CCCCTCTCTT TTACTCTGCT	1380
	CAAAAGCAT CTCTCTCCCC ATTACCCAAC CTTGGTCATA AGTGTGCCTG GCTGGTTGTC	1440
60	AGATATTGATGTC TCTGCTTTGT AAAAATTGGC CATTAGTGCA TTTATTGAGA TGATCTCTAA	1500

	AGAGCTATGC CCTGACCTAC CCCTGATTCT ATGACATTGG GGCCCTTCTT TTGCTGAAAC	1560
5	TGCCCTACGT AATGGTTTA CTCCCTGAAA GAGATTGAC CGAACCCATT TTATGCCAAG	1620
	TGCTGCCCTG CACTGTTCT GCAATATGTG GTGTATGCTG TGGTGATCTT GCTGGGAATG	1680
	ATTATAAGTG TGTGTGTGGT GGGGGAGTGG GTATTACATG CATTGCTGAA GAGTCAAAAA	1740
10	AAAAAAAAAA AAACTCGA	1758

15 (2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 888 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

25	CTGTTAGAAT GCCCAGTTA CCTGGATGGC AACCCAACAG TGCTCCTGCC CACCTGCC	60
	TCAATCCTCC TAGAATTCAAG CCCCAATTG CCCAGTTACC AATAAAAATC TGTACACCAG	120
30	CCCCAGGGAC AGTCTCAAAT GCAAATCCAC AGAGTGASMC ACCACCTCGG GTAGAATTG	180
	ATGACAACAA TCCCTTAGT GAAAGTTTC AAGAACGGGA ACGTAAGGAA CGTTTACGAG	240
	AACACCAAGA GAGACAAACGG ATCCAACCTCA TGCAGGAGGT AGATAGACAA AGAGCTTTGC	300
35	ACCAGAGGAT GGAAATGGAG CAGCATGGTA TGGTGGGCTC TGAGATAAGT AGTAGTAGGA	360
	CATCTGTGTC CCAGATTCCC TTCTACAGTT CCGACTTACC TTGTGATTT ATGCAACCTC	420
40	TAGGACCCCT TCAGCAGTCT CCACAAACACC AACAGCAAAT GGGCAGGTT TTACAGCACC	480
	AGAATATACA ACAAGGATCA ATTAAATTCAAC CCTCCACCCA AACTTTCATG CAGACTAATG	540
	ACCGAGGCCAG GTAGGCCCTC CTTCATTTGT TCCTGATTC CAATCAATCC CTGTTGGAAG	600
45	CCCAAATTTT TCTTCTGTGA AGCAGGGACA TGGAAATCTT TCTGGGACCA GCTTCCAGCA	660
	GTCCCCAGTG AGGCCTCTT TTACACCTGC TTTACCAGCA GCACCTCCAG TAGCTAATAG	720
50	CAGTCTCCCA TGTGGCCAAG ATTCTACTAT AACCCATGGA CACAGTTATC CGGGATCAAC	780
	CCAATCGCTC ATTCACTTGT ATTCTGATAT AATCCCAGAG GAAAAAGGNN AAAAAAAARA	840
	AAAARAAAARA ARAAAGGAGA TGATGATGCA GAATTCCACC AAGGCTCC	888

55

(2) INFORMATION FOR SEQ ID NO: 175:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2379 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

	GGCAGAGCTA GTGTGGACTC CATCCCCCTG GAGTGGGATC ACGNCTATGA CCTCAGTCGG	60
10	GACCTGGAGT CTGCAATGTC CAGAGCTCTG CCCTCTGAGG ATGAAGAAGG TCAGGATGAC	120
	AAAGATTTCT ACCTCCGGGG AGCTGTTGSC TTATCAGGGG ACCACAGTGC CCTAGAGTCA	180
15	CAGATCCGAC AACTGGCAA AGCCTGGATG ATAGCCGCTT TCAGATAACAG CAAACCGAAA	240
	ATATCATTG CAGCAAAACT CCCACGGGGC CGGAGCTAGA CACCAGCTAC AAAGGCTACA	300
	TGAAACTGCT GGGCGAATGC AGTAGGAGTA TAGACTCCGT GAAGAGACTG GAGCACAAAC	360
20	TGAAGGAGGA AGAGGAGAGC CTTCCCTGGCT TTGTTAACCT GCATAGTACC GAAACCCAAA	420
	CGGCTGGTGT GATTGACCGA TGGGAGCTTC TCCAGGCCA GGCATTGAGC AAGGAGTTGA	480
25	GGATGAAGCA GAACCTCCAG AAGTGGCAGC AGTTAACTC AGACTTGAAAC AGCATCTGGG	540
	CCTGGCTGGG GGACACGGAG GAGGAGTTGG AACAGCTCCA GCGCTGGAA CTCAGCACTG	600
	ACATCCAGAC CATCGAGCTC CAGATCAAAA AGCTCAAGGA GCTCCAGAAA GCTGTGGACC	660
30	ACCGCAAAGC CATCATCCTC TCCATCAATC TCTGCAGCCC TGAGTTCAAC CAGGCTGACA	720
	GCAAGGAGAG CGGGACCTG CAGGATCGCT TGTSGCAGAT GAATGGGCGC TGGGACCGAG	780
35	TGTGCTCTCT GCTGGAGGAG TGGGGGGGCC TGCTGCAGGA TGCCTGATG CAGTGCCAGG	840
	GTTTCCATGA AATGAGCCAT GGTTTGCTTC TTATGCTGGA GAACATGAC AGAAGGAAAA	900
	ATGAAATTGT CCTTATTGAT TCTAACCTTG ATGCAGAGAT ACTTCAGGAC CATCACAAAC	960
40	AGCTTATGCA AATAAAGCAT GACCTGTTGG AATCCCAACT CAGAGTAGCC TCTTGCAAG	1020
	ACATGTCTTG CCAACTACTG GTGAATGCTG AAGGAACAGA CTGTTTAGAA GCCAAAGAAA	1080
45	AAGTCCATGT TATTGAAAT CGGCTCAAAC TTCTCTGAA GGAGGTCAGT CGTCATATCA	1140
	AGGAACATGGA GAAGTTATTA GACGTGTCAA GTAGTCAGCA GGATTTGTCT TCCCTGGTCTT	1200
	CTGCTGATGA ACTGGACACC TCAGGGCTTG TGAGTCCAY ATCAGGAAGG AGCACCCCAA	1260
50	ACAGACAGAA AACGCCACGA GGCAAGTGTGTA GTCTCTCACA GCCTGGACCC TCTGTCAAGCA	1320
	GTCCACATAG CAGGTCCACA AAAGGTGGCT CCGATTCCCTC CCTTTCTGAG CCARGGCCAG	1380
55	GTGGTCCGG CGCGGGCTTC CTGTTAGAG TCCTCCGAGC AGCTCTTCCC CTTCAGCTTC	1440
	TCCTGCTCTT CCTCATCGGG CTTGCCTGCC TTGTACCAAT GTCAAGAGGAA GACTACAGCT	1500
	GTGCCCTCTC CAACAACCTT GCCCGGTCAAT TCCACCCAT GCTCAGATAAC ACGAATGGCC	1560
60	CTCCTCCACT CTGAACTAAG CAGATGCCAT CTGCAGAAGT GCTGGTAGCA TAAGGAGGAT	1620

	CGGGTCATAA GCAATCCCAA ACTACCAACA AGAGGACCTT GATCTTGGCG AAAGCCMTCG	1680
5	GTTGTGGCAGC TTTAGCCTCC TCCAGATCAC ATGTGTGCAA ATTATGGCTT CAGAGGTGGA	1740
	AGATAAACAG TGACGGGGGA ACAAACAGAC AACAAAGAAGG TTTGGAAGAA ATCTGGTTG	1800
	AGACTCTGAA CCTTAGCACT AAGGAGATTG AGTAAGGACC TCCAAAGTTC CCCGGACTCA	1860
10	TGAATTCTGG GCCCTTGGCC NATTCTGTGC ACAGCCAAGG ACTTCAGTAG ACCATCTGGG	1920
	CAGCTTTCCC ATGGTGCTGC TCCAACCATC AGATAAATGA CCCTCCCAAG CACCATGTCA	1980
15	GTGTCGTACA ATCTACCAAC CAACCAGTGC TGAAGAGATT TTAGAACCTT GTAACATACA	2040
	ATTTTTAAGA GCTTATATGG CAGCTTCCTT TTTACCTTGT TTTCTTTGG GGCATGATGT	2100
	TTAACCTTT GCTTTAGAAG CACAAGCTGT AAACTAAAAA GGCACTTTTT TTTAGAGGTA	2160
20	TAAAGAAAAA CTAGATGTAA TAAATAAGAT CATGGAAGGC TTTATGTGAA AAAAGTTGAA	2220
	TGTTATAGTA AAAAAAAAAG ATATTTATGT ATGTACAGTT TGCTAAAGCC AAGTTTTGTT	2280
25	TGTATTGATT TCTTGCATT TATTATAGAT ATTATAAAAT AAAAAAAAAA AAAAAAAAC	2340
	TCGAGGGGGG GCCCGGTACC CAATTGGCCC TATAGTGAG	2379

30

(2) INFORMATION FOR SEQ ID NO: 176:

35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1348 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:	
	GCGCCTTCAC GATGCCGGCG GTCACTGGTC CAGGTCCCTT ATTCTGCCTT CTCCCTCTGC	60
	TCCTGGACCC CCACAGCCCT GAGACGGGT GTCCCTCTCT ACGCAGGTTT GAGTACAAGC	120
45	TCAGCTTCAA AGGCCAAGG CTGGCATTGC CTGGGGCTGG AATACCCTTC TGGAGCCATC	180
	ATGGAGGTGA GGGCCAGGGG TGGGGACCGC TATGCCAGG GTCCCTCAA GTCCTGGAGG	240
50	GGCTGTRACT TGGTGGGGAG TGGTCTGTGTC ACAGCCATCC TCTGTCCAGG GTGGGGCAAG	300
	GCCTGGGACA GTGCCAGGCA CCCCAGGACC CCTTCCAGGC TTGTCTCTG CTCCACCGCC	360
	TCAACACCCCC CCACCCCTGC CCAAGCTGTT TCTCTCTGTC CTCTCTNNNTT CCCCTGCCCCA	420
55	GGACTTCTCT CTTCTCTCT GCCTCTCCCTT GGACCCCTGC CCTTCCCTCTA CCTCTGACCT	480
	GTGAACACAC AGACACATGC TCACACACTA AGTCCCARGC ACACMSAAAG GCAATGTGGA	540
60	CCAGCACAAA CCTCCACTCT CCCGGCTCCA TCCCACGGG CCTGTGGCTG CCCATGAAA	600

	CTGGGGGCTA CCTGGAGGGA AGCATCCTCA TCCCAGGTGA GTGGCACCA GCCCTTCCCT	660
	GTATGTGTGT TGTGGGTGGA ACCAGGCATG AGACCATCTT AGCCCATAGG TTTGTATTCA	720
5	GGGACTTCCA AACCCAGACC TACAAAGAGT GTGTCTTCTA CCAGATCTTG TTCAAAAAAG	780
	GGTTTGTGAT GATGGAACTA CACGATAGAG GGAGTGAGCA AGAACATGA GGATTAGAGT	840
10	GGAGCGTGAA ATAGTCTAGG ACCATGGCTT CCAAAACATA TGCTGTGAGG TCTGTCCACC	900
	TGAGAGTTGG CCCATGGATT TAATTCTGAG CCTCTTAGCA GGCAAAGCAA AGACAGAAAAG	960
	CAGATCGGCT GTGGATTCT GTCTATAAAA TGTGAGTTCT TGGCCGGGTG CGGTGGCTCA	1020
15	CGCCTGTAAT CCCGGCGCTT TGGGAGGCCA GGGCGGATGG GTGCGGAGGT CAGGAGGTG	1080
	GAAACCATCC TGGCCGGAAT GGTGAAGCCC TGACTCTACT AGAACGTGCAA AGATTGGCTG	1140
20	GGTGTGGTGG CGTGCCTGCTG TGGTCCCAGC TTCTCGGGAG GCTGAGGCGG GAGAGTTGCT	1200
	TGGGCCTGGG AGGCGGAGGT TGCGGTGAGC TGAGATCTG CCATTGCACT TCAGCCTGGG	1260
	CACAGAGCCA GACTCTGGCT CAAAAAAA AAAAAAAA ACTCGAGGGG GGCCCGTACC	1320
25	CAATTGCCG NATATGATCG TAAACAAT	1348

30 (2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1502 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

40	CTCAAAATAA ATAAATAAAT AAAAATTGT ATTCCATTGA TTTGGGTAGA CACCAGGAAT	60
	GTCATTTCT AACAAACCTTT CCAGGCGATC CTATAGTAAG TCATCTGTGG ACTACTTTAA	120
45	GAAACTCTTC TATAGAGAAAT GGAGTTGGAT TAATAATAGG TGATTTTTTA CACTGGACTG	180
	ATTCACAAGA ACCTAAACAG TAGTCCATGA AGCTGCTCAT CTGTGGTAAC TATTTGGCCC	240
	CGTCTCACTC TGAAAGCAGC AGGAGATGTT GTTTACTTTG TTTCTATCCC CTTTGTCTGG	300
50	AGATTAATTG TCGAATGAAA GTTTTCTCT CTATGCCATT CCTGGTTCTT TTCCAAAGCC	360
	TCATACAAGA GGATTAGTC ACAATGCATG CATTACCTTT TAAAAGAATG CGATATTGAT	420
55	ACCGATGCTT ACTTTTTTTT TTCTTNACTA CTTGTTTAT TCCTTCCAGN AAAGTATAGC	480
	CCGCCTTCT ATAGCATAGT TCTCTTCTAGG TCGAATGATT CCTATAAGAT TTCTCATTAT	540
	TAAATCATGC ATTTTCAAG ATGGAATCAA TMTTTGATTT AATCTAAGCT GATATTCTCA	600
60	TTTGTAGAA GAACAAACCTA CATGCTAGAG AGAGAGGAGG AAATATAACCC ACGACCACAC	660

	AGCCAGTTAG TATCCAGTTG GTGCTGGACT CCAGCCAGGT GTCCTGCCTC ATGGTAGTTA	720
5	AATGATATAT AGAAAAGGTA AATTTTAAAGA GAAATATTTA TTAATATAATT CCTATAAAAC	780
	ATTTTAAAGG TAACCACATA AAAATGGTTA ATTTTTCCAT TCCAAAGTAA ATGCTAAGCA	840
	TGTTTATTAA TGAAGCAGTA CTTCTGATTA GTATATGACA TTCTGAAGTT AATTAACIC	900
10	ATTGCACCAA ATGTGTCTTC CTTGGTATAG TGGAGGATTT GAGGATTGGA ATATAGAGTA	960
	GAGTGCTTGC TTAAGCCTGG GAGCCCACATCT TTATAGCTAT TTGATGTAAG AAAAGAGACA	1020
15	TGGNCCATTCTAAACTATA TAAGGTGAGT GTGTCTATTCC CACCGAGATA TAAAGGAAAA	1080
	ACGAAACCTTT TTTGATTCCC ACCTTCCCCAG CCTCACCTAG CCATCTTCCA GCCTCAAATA	1140
	TAGAGATGTT AGTGCAAGGT CCTGGGCTCT AGGTGATCAT TTCATAAGTC CTTTACAGAT	1200
20	AAAGAAAAAG TAGTGTGTTGT ATGTTGTTT TTAAGTAACC CCAAAACAAA TTTATATTGT	1260
	ATTCAGCAA ATTGGAATTG AGGTGTTAA TTTTAAACCA TGAAGTGCCT GCTGTTTAA	1320
25	GCATTGACTT GTATAAAAG AATTCATGT CTCCAGTAAG CTTATGGGTT TTCTCATTTC	1380
	TAGGTATATG GCTTTAACATC ATGTAAGTG AAACATTAGT TTTCTTGAT TTTATTACAG	1440
	GTTCCTTGTGTT GCAATAAAGA TGCTGCTGAA ATTAATTGAA AAAAAAAA AAAAAAAACTC	1500
30	GA	1502

35 (2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1637 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

45	ATTTCTAGC CCACAAGGAC TGAAGTTCAAG ATCCAAAAGT TCACCTTGCTA ATTATCTTCA	60
	CAAAATGGA GAGACTTCTC TTAAGCCAGA AGATTTGAT TTTACTGTAC TTTCTAAAAG	120
50	GGGTATCAAG TCAAGATATA AAGACTGCAG CATGGCAGCC CTGACATCCC ATCTACAAAA	180
	CCAAAGTAAC AATTCAAACG GGAACCTCAG GACCCGAAGC AAGTGCAAAA AGGATGTGTT	240
	TATGCCGCCA AGTAGTAGTT CAGAGTTGCA GGAGAGCAGA CGACTCTCTA ACTTTACTTC	300
55	CACTCATTG CTTTGAAAG AAGATGAGGG TGTTGATGAT GTTAACCTCA GAAAGGTAG	360
	AAAGCCAAA GGAAAGGTGA CTATTTGAA AGGAATCCCA ATTAAGAAAA CTAAAAAAGG	420
60	ATGTAGGAAG AGCTGTTCAAG GTTTGTTCM AAGTGATAGC AAAAGAGAAT CTGTGTGAA	480

430

	TAAAGCAGAT GCTGAAAGTG AACCTGTTGC ACAAAAAGT CAGCTTGATA GAACGTGTCIG	540
	CATTTCATGAT GCTGGAGCAT GTGGTGAGAC CCTCAGTGTG ACCAGTGAAG AAAACAGCCT	600
5	TGTAAAAAAA AAAGAAAGAT CATTGAGTTC AGGATCAAAT TTTTGTCTG AACAAAAAAC	660
	TTCTGGCATC ATAAACAAAT TTGTTCAAGC CAAAGACTCA GAACACAACG AGAAGTATGA	720
10	GGATACCTTT TTAGAATCTG AAGAAATCGG AACAAAAGTA GAAGTTGTGG AAAGGAAAGA	780
	ACATTTGCAT ACTGACATTT TAAAACGTGG CTCTGAAAATG GACAACAACG GCTCACCAAC	840
	CAGGAAAGAC TTCACTGAAG ATACCATCCC ACGGAACACA GATAGAAAGA AGGAAAACAA	900
15	GCCTGTATTT TTCCAGCAA TATAACAAAG AAGCTCTTAG CCCCCCACGA CGTAAAGCCT	960
	TTAAGAAATG GACACCTCCT CGGTACCTT TTAATCTCGT TCAAGAAACA CTTTTTCATG	1020
20	ATCCATGGAA GCTTCTCATC GCTACTATAT TTCTCAATCG GACCTCAGGC AAAATGGCAA	1080
	TACCTGTGCT TTGGAAGTTT CTGGAGAAGT ATCCCTCAGC TGAGGTAGCA AGAACCGCAG	1140
	ACTGGAGAGA TGTGTCAAGA CTTCTTAAAC CTCTTGGCT CTACGATCTT CGGGAAAAAA	1200
25	CCATTGTCAA GTTCTCAGAT GAATACCTGA CAAAGCAGTG GAAGTATCCA ATTGAGCTTC	1260
	ATGGGATTGG TGCACCCCTGA AGACCACAAA TTAAATAAAAT ATCATGACTG GCTTTGGGAA	1320
30	AATCATGAAA AATTAAGTCT ATCTTAAACT CTGCAGCTTT CAAGCTCATC TGTTATGCAT	1380
	AGCTTGCAC TTCAAAAAAG CTTAATTAAG TACAACCAAC CACCTTCCA GCCATAGAGA	1440
	TTTTAATTAG CCCAACTAGA AGCCTAGTGT GTGTGCTTTC TTAATGTGTG TGCCAATGGT	1500
35	GGATCTTGC TACTGAATGT GTTGAAACAT GTTTGAGAT TTTTTAAAAA TAAATTATTA	1560
	TTTGACAACA ATCCAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1620
40	AAAAAAAAAA AAAAAAAA	1637

(2) INFORMATION FOR SEQ ID NO: 179:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2911 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

55	GGTGGTTTTT GTTCTGCAAT AGCGGGCTTA GAGGGAGGGG CTTTTTCGCC TATACCTACT	60
	GTAGCTTCTC CACGTATGGA CCCTAAAGGC TACTGCTGCT ACTACGGGGC TAGACAGTTA	120
	CTGTCTCAGC TCTAGGATGT GCGTTCTTCC ACTAGAAGCT CTTCTGAGGG AGGTAATTAA	180
60	AAAACAGTGG AATGGAAAAAA CAGTGCTGTA GTCATCCTGT AATATGCTCC TTGTCAACAA	240

	TGTATACATT CCTGCTAGGT GCCATATTCA TTGCTTTAAG CTCAAGTCGC ATCTTACTAG	300
5	TGAAGTATTG TGCCAATGAA GAAAACAAGT ATGATTATCT TCCAACTACT GTGAATGTGT	360
	GCTCAGAACT GGTGAAGCTA GTTTCTGTG TGCTTGTGTC ATTCTGTGTT ATAAAGAAAG	420
	ATCATCAAAG TAGAAATTG AAATATGCTT CCTGGAAGGA ATTCTCTGAT TTCATGAAGT	480
10	GGTCCATTCC TGCCTTCTT TATTCTCTGG ATAACCTGAT TGTCTCTAT GTCCTGTCC	540
	ATCTCAACC AGCCATGGCT GTTATCTCT CAAATTTAG CATTATAACA ACAGCTCTTC	600
15	TATTCAGGAT AGTGCTGAAG ANCGTCTAA ACTGGATCCA GTGGCTTCC CTCCTGACTT	660
	TATTTTGTC TATTGTGCC TTGACTGCCG GGACTAAAAC TTTACAGCAC AACTTGGCAG	720
	GACGTGGATT TCATCAGGAT GCCTTTTCA GCCCTTCCAA TTCCCTGCTT CTTTCAGAA	780
20	ATGAGTGTCC CAGAAAAGAC AATTGTACAG CAAAGGAATG GACTTTCTT GAAGCTAAAT	840
	GGAAACACCAC AGCCAGAGTT TTCAGTCACA TCCGTCTGG CATGGCCAT GTTCTTATTA	900
	TAGTCCAGTG TTTTATTCT TCAATGGCTA ATATCTATAA TGAAAAGATA CTGAAGGAAG	960
25	GGAACCAGCT CACTGAARGC ATCTTCATAC AGAACAGCAA ACTCTATTTC TTTGGCATTC	1020
	TGTTTAATGG GCTGACTCTG GGCTTCAGA GGAGTAACCG TGATCAGATT AAGAACTGTG	1080
30	GATTTTTTA TGGCCACAGT GCATTTTCAG TAGCCCTTAT TTTTGTAACT GCATTCCAGG	1140
	GCCTTTCAGT GGCTTCATT CTGAAGTTCC TGGATAACAT GTTCCATGTC TTGATGGCCC	1200
35	AGGTTACAC TGTCATTATC ACAACAGTGT CTGCTCTGGT CTTTGACTTC AGGCCCTCCC	1260
	TGGAATTTT CTTGGAGCC CCATCAGTCC TTCTCTCTAT ATTATTTAT AATGCCAGCA	1320
	AGCCTCAAGT TCCGGAATAC GCACCTAGGC AAGAAAGGAT CCGAGATCTA AGTGGCAATC	1380
40	TTTGGGAGCG TTCCAGTGGG GATGGAGAAG AACTAGAAAG ACTTACCAAA CCCAAGAGTG	1440
	ATGAGTCAGA TGAAGATACT TTCTAACTGG TACCCACATA GTTGCAGCT CTCTTGAAACC	1500
	TTATTTTCAC ATTTTCAGTG TTGTAAATAT TTATCTTTTC ACTTTGATAA ACCAGAAATG	1560
45	TTTCTAAATC CTAATATTCT TTGCTATAT CTAGCTACTC CCTAAATGGT TCCATCCAAG	1620
	GCTTAGAGTA CCCAAAGGCT AAGAAATTCT AAAGAACTGA TACAGGAGTA ACAATATGAA	1680
50	GAATTCACTTA ATATCTCACT ACTTGATAAA TCAGAAAGTT ATATGTGCAG ATTATTTCC	1740
	TTGGCCTTCA AGCTTCCAAA AAACCTGTAA TAATCATGTT AGCTATAGCT TGTATATACA	1800
	CATAGAGATC AATTGCCAA ATATTCACAA TCATGTAGTT CTAGTTACA TGCCAAAGTC	1860
55	TTCCCTTTT AACATTATAA AAGCTAGGTT GTCTCTGAA TTTGAGGCC CTAGAGATAG	1920
	TCATTTTGCA AGTAAAGAGC AACGGGACCC TTTCTAAAAA CGTTGGTTGA AGGACCTAAA	1980
60	TACCTGGCCA TACCATAGAT TTGGGATGAT GTAGTCIGTG CTAAATATTT TGCTGAAGAA	2040

	GCAGTTTCTC AGACACAACA TCTCAGAATT TTAATTTTA GAAATTCAIG GGAAATTGGA	2100
5	TTTTGTAAAT AATCTTTGA TGTTTAAAC ATTGGTCCC TAGTCACCAT AGTTACCACT	2160
	TGTATTTAA GTCAATTAAA CAAGCCACGG TGGGGCTTT TTCTCCTCAG TTGAGGAGA	2220
	AAAATCTTGA TGTCATTACT CCTGAATTAT TACATTITGG AGAATAAGAG GGCAATTAT	2280
10	TTTATTAGTT ACTAATTCAA GCTGTGACTA TTGTATATCT TTCCAAGAGT TGAAATGCTG	2340
	GCTTCAGAAT CATAACCAGAT TGTCAGTGAA GCTGATGCCT AGGAACCTTT AAAGGGATCC	2400
	TTTCAAAAGG ATCACTTAGC AAACACATGT TGACTTTAA CTGATGTATG AATATTAATA	2460
15	CTCTAAAAAT AGAAAGACCA GTAATATATA AGTCACTTA CAGTGCTACT TCACACTTAA	2520
	AAGTGCATGG TATTTTICAT GGTATTTGC ATGCAGCCAG TTAACTCTCG TAGATAGAGA	2580
20	AGTCAGGTGA TAGATGATAT TAAAAATTAG CAAACAAAAG TGACTTGCTC AGGGTCATGC	2640
	AGCTGGGTGA TGATAGAAGA GTGGGCTTTA ACTGGCAGGC CTGTATGTTT ACAGACTACC	2700
	ATACTGTAAA TATGAGCTTT ATGGTGTCA TCTCAGAAC TTATACATTT CTGCTCTCCT	2760
25	TTCTCCTAAG TTTCATGCAG ATGAATATAA GGTAATATAC TATTATATAA TTCATTGTCG	2820
	ATATCCACAA TAATATGACT GGCAAGAATT GGTGGAAATT TGTAATTAAA ATAATTATTA	2880
30	AACCTAAAAA AAAAAAAA AAAAAGTCGA G	2911

35 (2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 519 base pairs
- (B) TYPE: nucleic acid
- 40 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

45	GGCACGAGCC CCAGGCCAGC CAGGGCCAGG CCTACTTTGG CCACCCCTAA ATTAGAATGT	60
	GGGGTCAGGG GTCACAGAAA ACCCATTTCT CTGACCTAGT GTTGGGCGTC CGGGAACTCT	120
50	GTGCCCAACC TTCAGACCCCT GGCACTCTC ACTGAGGCCA TTGGCCAGA GCGCGCCATC	180
	CCCCGARACC CCCGGGAGCC GCCTGTTGCC ACGTCCACAC CTGCCACACC CTCTGCCGGG	240
	CCCCAGCCCC TCCCAACCGG GACCGTGCTG GTCCCTGGGG GTCCCTGCCCTT ACCTTGCTTT	300
55	GGGGAGGCAT GGGCCCTCCT CCTCCCCACCC TGCCGGCCGT CACTCACCTC TTGCTTCCTGG	360
	TCCCCCAGGC CTAGCCCTTG GAAGGGAGACA GGAGTCTAGG GAGGCTGAAG CCCACTCCCG	420
60	GGGAGGCCCG TGCTCCCTCCA GCCCCAGGGA CAGCAAGGAA AAGAGAAGAG AGCAGAGCAT	480

TTCATGGCTC TAATAAAAAA AAAAAAAA AAAACTCGA

519

5

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 968 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

15	TCCCCCTGGG GCGGGAAAAA GCGGGGGTGG CCTGNCCATT GGTINTCCAT GCCGCCCCGCC	60
	CATGCCCGAG TACTAGCCTG CAGTCCCAAT GTAGCCCCCTC CCTCYTCCMA GAGCCCYTCM	120
20	AACCGCCCCG STCANTIGTG ATITCAGGAG GATTGATGA AGATGTTAAA GCGAAAGTGG	180
	AGAACCTTCT CGGGATTTC AGCCTGGAAA AAACGGACCC TGTTAGGCAA GCACCCCTGCA	240
25	GCCCTCCCTG TCCCCTCTT CCCCTCCCT TCYCCCCGCC GTGGAGACAG CTGTTYTCA	300
	CAGGGCTCTC CGCAGGGAGG GGGCCGGCTC CTTCCCTGGC AGAACACATCC TTGCCCCTTGT	360
	CACACAAGTC AGCCTCCATC TGGCAGCTC TGTGGATGCG CTGCTGGAGG GCAACAGGTA	420
30	TGTCACTGGC TGGTTCAGCC CCTACCACCG CCAGCGGAAG CTCATCCACC CGGTCAATGGT	480
	TCAGCACATC CAGCCCGCAG CGCTCAGCCT CCTGGCACAG TGGAGCACCC TCGTGCAGGA	540
35	GCTGGAGGCT GCCCTGCAGC TGGCTTTCTA CCCGGATGCC GTGGAGGAGT GGCTGGAGGA	600
	AAACGTGCAC CCCAGCTGC AGCGGCTGCA ARCTCTGCTG CAGGACCTCA GCGAGGTGTC	660
	TGCCCCCCCG CTGCCACCCA CCAGCCCTGG CAGGGACGTT GCTCAGGACC CCTGAGGGGA	720
40	GAGCTCATGC CAGGGGCTC CTGCTGGAGG CTGGGGGGC TCTGCWYTKY CWWWTGGCCT	780
	GGGCAATACG GCCCACGTGG GCGTCGTGCC CTCTGGCCCA GCAGTGTCTT GCCCACACTC	840
45	AGTTCTTGAG GGCCCTGGGC AGCCCCCTGGG GGAGAGACTA GAAAACACAG AAGGAAGCAG	900
	CACAGGGAGA CCCGCTTTGT GATCTGCCATG TGTGACACTG ATTCTTTGGA AATAAAGAGT	960
	GGAAGCTG	968

50

(2) INFORMATION FOR SEQ ID NO: 182:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

	TGTAAAAGTT ATCAGTAATC CTAATTCTTT TCCTGGGTTT TCCTTTTGTC ACTTATTAAT	60
5	CAGTTTTGGA AAGGACGAAT GAATTTAGAG ATGTACTCTG GAGCAGTATC ATGTTAAACC	120
	AGGGGTATAT TAGAAAAATC ATCCTCATAA TCATTCTGGG AAGTTTTCC TCCCCAAAAA	180
10	AAGCCATCCT GATGGGTTTT CAAAACCAGA AAAAGCTCT TAATGAGGAA CAGACCACTG	240
	GAATACCCAT GAGCATCTCA GGAAAACCTGA GACCCCTCGAG AACCCCTGAT TTCTGTGCAAC	300
	CCCCAAGGTT TCAGAGCCAG CAGCCCAGTG CTGTGGTTGA CAGACGTGGT TTTKTGGRGA	360
15	AAGCAGCCAG AGGCCAGGAA TTTTCAGAGT CGTGAGTCAC GRTYTCCCAC CCAAGATTAG	420
	AGCAMAGATT AGCCATACTG AGATTGGTA AAATCATTCT GTCTAAGCAA TGGAGGTGTG	480
20	TGCAMACGTG CAGTGCCTGT TCACAGGGGA TGCAGGCAGA TCSYGGGTTT AGGATGGGR	540
	AGGCCACCGC ACCCCCCYTC AYTGCTCTGC ACCTGCTCCC TCACGTGGAC ACTGTCCACA	600
	ACTGTGGCTC TCACAGGACA GTTGCCCAAG GAGCTCATAT CTTATTGGAG ATAGGGGTC	660
25	GTACAGGTGA CATTCACTGAG CAGTGTGAGC CGGGTGACAT GGGGGTGTCA ACCCAGCATC	720
	TGTCCAGGAG CTCCCTCTGC AGCGGCTCTG GCAGGTGGCC TGAGGCTCCT TTTTGAGAGA	780
30	GAACTGTTTG GCCTTCCTGT CTCCCTCTCCT CTGATCTGTT CTTTCTTGGAA ACACCACCA	840
	AGAACGTCAC CTCCCTCAC AGATTGTGAG CTCCCTGGAGG GCAGGAGCTG TGTCCCTCTA	900
	TTCATCTTCC TATCCCCAGA ACCTTGCACA GATCCTGGAA TGTGGTAGGT GCTCAGTAAA	960
35	TGTGTGTTGA ATAAATGAAT GAATGAATGA ACAAAATGAAT GAATTGGCTT ACTTCAAGGC	1020
	AAAAGAACCA TGAAACTGTA TTTTGAGTTT CTATGTTATA GCAGTCAGCA AATCCTATTA	1080
	AATACTTTGT GTTCCAAGC AAAAAAAAAA AAAAAAAAAA AAAACTCGA	1128
40		

(2) INFORMATION FOR SEQ ID NO: 183:

45	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 2276 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
50	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

55	CGCGGGCGTC TGACCTCATG GCGTAGAGCC TAGAACACGC GCAGGCTCCC AGCGGAGTCC	60
	GTTATGGCCCG CTGCCGTCCC GAAGAGGATG AGGGGGCCAG CACAAGCGAA ACTGCTGCC	120
	GGGTGGGCCA TCCAAGCCCT TGTGGGGTTG GCGCGGCCGC TGGCTCTGGC GCTCCTGCTT	180
60	GTGTCCGCCG CTCTATCCAG TGTGTATCA CGGACTGATT CACCGAGCCC AACCGTACTC	240

	AACTCACATA TTTCTACCCC AAATGTGAAT GCTTTAACAC ATGAAAACCA AACCAAACCT	300
5	TCTATTTCGG AAATCAGCAC CACCCCTCCCT CCCACGACGA GTACCAAGAA AAGTGGAGGA	360
	GCATCTGTGG TCCCTCATCC CTGGCCTACT CCTCTGTCTC AAGAGGAAGC TGATAACAAT	420
	GAAGATCCTA GTATAGAGGA GGAGGATCTT CTCATGCTGA ACAGTTCTCC ATCCACAGCC	480
10	AAAGACACTC TAGACAATGG CGATTATGGA GAACCAGACT ATGACTGGAC CACGGGCC	540
	AGGGACGACG ACGAGTCTGA TGACACCTTG GAAGAAAACA GGGGTTACAT CGAAATTGAA	600
15	CAGTCAGTGA AATCTTTAA GATGCCATCC TCAAATATAG AAGAGGAAGA CAGCCATTTC	660
	TTTTTCATC TTATTATTTTG TGCTTTTGCG ATTGCTGTG TTTACATTAC ATATCACAAC	720
	AAAAGGAAGA TTTTCTCT GGTCAAAGC AGGAAATGGC GTGATGGCCT TTGTTCCAAA	780
20	ACAGTGAAT ACCATGCCCT AGATCAGAAT GTTAATGAGG CAATGCCCTC TTTGAAGATT	840
	ACCAATGATT ATATTTTTA AAGCACTGTG ATTGAAATTG GCTTATGTAA TTTTATTGCG	900
	TTGACTTTTT ATATGATAATT GTGCAAATGT TTGCCATAGG CAATTGGTAC TTAAATGAGA	960
25	GGTGAGTCTC TCCTTGCCT TGGTGCTTTG GAAATTAAAT GTCACAAACG AGTATATAAT	1020
	TTTTATCTG TACTTTAGA GCTGAGTTTA ATCAGGTGTC CAAAATGTGA GTAAACATT	1080
30	ACCTTATATT TACACTGTTA GTTTTATTG TTTTAGATT ATTATGCTTC TTCTGGAAGT	1140
	ATTAGTGATG CTACTTTAA AAGATCCCAA ACTTGAACT AAATTCTGAC ATATCTGTTA	1200
	CTGCTGACTC ACATTCAATT CCGCCATTCA AAATACTATT TTTTATCCAC ATTTTTTTT	1260
35	GTTCCAAAC TGTAATGTAC AAGGATATGT GTGATAATGC TTTGGATTG AGTAATATT	1320
	TTTTTCTTC CAAGAAAATC GCTTGGATA TTTTAGATA ATTAAACAT ATTAGGAT	1380
40	AATGATATTG CTCAATCTGA CCACAATTG AGGTAAAACA TTAAATGTGT CAGAAATCTT	1440
	GGCAACAGAG ACTCTGCAGC TTGCAGTGGG CATAGATAAA ATGTTACAGA GATACTATT	1500
	TTTGGTTGG AATTACTATA TTAAATTAG AAGCAGAAAC TGGTAAATG TTAAATACAT	1560
45	GTACAATTGC TTTAGTTAG CAATTGATTG TAGCATGGGT TCCTCCAAGG TTCAAGCAA	1620
	TGGGCAGAGT TTAAAATTAT ATCAGATTG TTTACTTCGT TTATTATTIT ACAGTAAATT	1680
50	TGAATAAAC TTAGGGTCA TTATCACTTA AATAACTTG TACCTAGGTC TTCAAATTA	1740
	AAATTATAACC TGAATGAAGT TGGTTGTATA CATAAAGGAT ATTGTGTAC AATTACCTT	1800
	TTTCCCCCAC ACTTGTTTC TTTGGTTTG TTTTTATGG CAACTGGAAA GTATTTACTA	1860
55	TGGGATTCAAT TTATGTCTGT CTTTCTATCA TAAAGAATTG ATCAATATGT AAATATGTGA	1920
	TTTGAACCAT GGTTGACTTA CAAGTGTAC TACAGCTTT TAGAAAACAT AGCCCTAATA	1980
60	TATGTTAACG AGGACCCGGG TGAGCCAGTG GGCTTGGCCT TTATGTAGAG CTGGAAGAAG	2040

5	GCCGTCCATC CTGTCCTTG GGCGGACAGT GTACTTCCT AATAGGGAAG GGAACCACAA	2100
	TGGAAATACC CCTGAACCGT TTTATTGCAG TAATTTTTT CATATCTGAA ACTATTATT	2160
	AATATTTGA ATAAGATTT AAAAATAAA TGGCAAAGAT ATAAATCTAA AAAAAAAA	2220
	AAAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAA	2276

10

(2) INFORMATION FOR SEQ ID NO: 184:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

25	TCCAAGCTAC GCCACTCGGG CTGGGGCGTT GGGAGCGGA GTGCAGAGCG TGTCGTGGC	60
	GGCGCCGGTG AGAACAGCGA GCGKAGGAG GGGTGCCAT GGCGGGCAG CAGTCCAGT	120
	ACGATGACAG TGGGAACACC TTCTCTACT TCCTCACCTC CTTCTGGGG CTCATCGTGA	180
	TCCCGGCAC ATACTACCTC TGGCCCCGAG ATCAGAATGC CGAGCAAATT CGATTAAGA	240
30	ATATCAGAAA AGTATATGGA AGGTGTATGT GGTACGTTA CGGTTATTAA AACCCCAGCC	300
	AAATATTATT CCTACAGTAA AGAAAATAGT TCTGCTTGCA GGATGGCAT TGTTCTTATT	360
35	CCTTGATAT AAAGTTCCA AAACAGACCG AGAATACCA AAATACAATC CTTATGAAGT	420
	ATTAATTTG GATCCTGGAG CCACAGTAGC AGAAATTAAA AAACAATATC GTTGCTGTC	480
	ACTTAAATAT CATCCAGATA AAGGAGGTGA TGAGGTTATG TTCATGAGGA TAGAAAAGC	540
40	TTATGCTGCT TTAACGGATG AAGAGTCCCG GAAAATTGG GAAGAATTG GAAATCCAGA	600
	TGGGCCTCAA GCCACAAGCT TTGGAATTGC CCTGCCAGCT TGGATAGTTG ACCAGAAAA	660
45	TTCAATTCTG GTTTTACTTG TATATGGATT GGCATTATG GTTATCCTTC CAGTTGTGT	720
	GGGCTCTGG TGGTATCGCT CAATACGCTA TAGTGGAGAC CAGATTCTAA TACGSACAAC	780
	ACAGATTAT ACATACTTIG TTTATAAAC CCGAAATATG GATATGAAAC GTCTTATCAT	840
50	GGTTTGGST GGAGCTCTG AATTGATCC TCAGTATAAT AAAGATGCCA CAAGCAGACC	900
	AACGGATAAT ATTCTAATAC CACAGCTAAT CAGAGAAATT GGCAGCATTA ATTAAAGAA	960
55	GAATGAGCCT CCACCTACCT GCCCATATAG CCTGAAGGCC AGAGTTCTTT TACTGTCTCA	1020
	TCTTGCTAGA ATGAAAATTC CTGAGACCT TGAAGAAGAT CAGCAATICA TGCTAAAAAA	1080
60	GTGTCTGCC CTACTTCAAG AAATGGTTAA TGTAATCTGC CAACTAATAG TAATGGCCCG	1140

	GAACCGTGAA GAAAGGGAGT TTCTGCTCC AACTTTGGCA TCCCTAGAAA ACTGCATGAA	1200
	GCTTTCTCAG ATGGCCGTTC AGGGACTTCA GCAATTAAAG TCTCCCTTC TGCGAGCTCCC	1260
5	TCATATTGAA GAGGACAATC TTAGACGGGT TTCTAATCAT AAGAAGTATA AAATTAAAAC	1320
	TATCCAGGAT TTGGTGAGTT TAAAAGAAC AGATCGTCAC ACTCTACTGC ACTTCCTTGA	1380
10	AGATGAAAAA TATGAAGAGG TTATGGCTGT CCTTGGGAGT TTTCATATG TGACCATGGA	1440
	TATAAAATCA CAGGTGTTAG ATGATGAAGA TAGCAACAAAC ATCACAGTAG GATCCTTAGT	1500
	TACAGTGTG GTTAAGTTGA CAAGGCAAAC AATGGCTGAA GTATTGAAA AGGACCAAGTC	1560
15	CATCTGTGCT GCAGAGGAAC AGCCAGCAGA AGATGGGCAAG GGTGAAACTA ACAAGAACAG	1620
	GACAAAAGGA GGATGGCAAC AGAAGAGTAA AGGACCCAAG AAAACTGCTA AATCAAAAAA	1680
20	AAAGAAACCT TTAAAAAAAAA AACCTACACC TGTGCTATT CCACAGTCAA AGCAACAGAA	1740
	ACAAAAGCAG GCAAATGGAG TCGTTGGAA TGAAGCTGCA GTAAAGGAAG ATGAAGAAGA	1800
	AGTTTCAGAT AAGGGCAGTG ATTCTGAAGA AGAAGAAACC AATAGAGATT CCCAAAGTGA	1860
25	GAAAGATGAT CGTACTGACA GAGACTCTGA TAGAGAGCAA GATGAAAAAC AAAACAAAGA	1920
	TGATGAAGCA GAGTGGCAAG AATTACAACA AAGCATAACAG CGAAAAGAGA GAGCTCTATT	1980
30	GGAAACCAAA TCAAAATAA CACATCCTGT GTATAGCCTT TACTTTCTG AGGAAAACA	2040
	AGAATGGTGG TGGCTTACA TTGAGATAG GAAGGAGCAG ACATTAATAT CCATGCCATA	2100
	TCATGTGTGT ACGCTGAAAG ATACAGAGGA GGTAGAGCTG AAGTTTCTG CACCAGGCAA	2160
35	GCCTGGAAAT TATCAGTATA CTGTTTCT GAGATCAGAC TCCTATATGG GTTGGATCA	2220
	GATTAACCA TTGGAAGTTC GGAAGTTCAT GAGGCTGAAG CCTGTGCCAG AAAATCACCC	2280
	ACAGTGGGAT ACAGCAATAG AGGGGGATGA AGACCAGGAG GACAGTGAGG GTTTGAAGA	2340
40	TAGCTTGAG GGAGGAAGAG GGAGGGAGGA ACCAACGGTGG TGGACTTAAG GCAGTTACTC	2400
	TGGAATGGGA CCCACAGTGT TTGACCAT ATTGAGCAA TTTTTTTTGC CCGTTTTTNG	2460
45	GAAGTGTGTTT CCNTNAANCC CAGGAACCAT TACAGAACCG	2500

50 (2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1337 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

60 CTTCCGGTTC TCCGGGCAGC TCCCCTGCT GTAGCTCTG CCACCTGCCA CGACGGGCC

60

	TCTCCCTGGC GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCCTCTTGGA	120
5	GCCAGCGTGG CGGGCCTGGC GGCTCCCGGG TGGTGAGAGA GCGGTCCGGG AACGATGAAG	180
	GCCTCGCAGT GCTGCTGCTG TCTCAGCCAC CTCTTGGCTT CCGTCCCTCT CCTGCTGTTG	240
	CTGCCCTGAAC TAAGCGGGYC CCTGGMAGTC CTGCTGCAGG CAGCCGAGGC CGGCCAGGT	300
10	CTTGGGCCTC CTGACCCCTAG ACCACGGACA TTACCGCCGC TGCCACOGGG CCCTACCCCT	360
	GCCCAGCAGC CGGGCCGTGG TCTGGCTGAA GCTGCGGGGC CGCGGGGCTC CGAGGGAGGC	420
15	AATGGCAGCA ACCCTGTGGC CGGGCTTGAG ACGGACGATC ACGGAGGGAA GGCCGGGGAA	480
	GGCTCGGTGG GTGGCGGCCT TGCTGTGAGC CCCAACCTG GCGACAAGCC CATGACCCAG	540
	CGGGCCCTGA CGGTGTTGAT GGTGGTGAGC GGCGCGGTGC TGGTGTACTT CGTGGTCAGG	600
20	ACGGTCAGGA TGAGAAGAAG AAACCGAAAG ACTAGGAGAT ATGGAGTTTT GGACACTAAC	660
	ATAGAAAATA TGGAAATTGAC ACCTTTAGAA CAGGATGATG AGGATGATGA CAACACGTTG	720
25	TTTGATGCCA ATCATCCTCG AAGATAAGAA TGTGCCTTTT GATGAAAGAA CTTTATCTTT	780
	CTACAATGAA GAGTGGAAATT TCTATGTTTA AGGAATAAGA AGCCACTATA TCAATGTTGG	840
	GGGGTATTAAAGTTACATA TATTTTAACA ACCTTTAAATT TGCTGTTGCA ATAAATACCG	900
30	TATCCTTTTA TTATATCTTT ATATGTATAG AAGTACTCTR TTAATGGCT CAGAGATGTT	960
	GGGGATAAAAG TATACTGTAA TAATTTATCT GTTTGAAAAT TACTATAAAA CGGTGTTTC	1020
35	TGATCGGTTT TGTGTTCTG CTTACCATAT GATTGTAAT TGTGTTATGT ATTAATCAGT	1080
	TAATGCTAAT TATTTTGCT GATGTCATAT GTAAAGAGC TATAAATTCC AACAAACCAAC	1140
	TGGTGTGTAA AAATAATTAA AAATTTCTTT TACTGAAAGG TATTCTCCAT TTTTGTGGGG	1200
40	AAAAGAAGCC AAATTTATTA CTTTGTGTTG GGGTTTTAA AATATTAAGA AATGTCTAAG	1260
	TTATTGTTTG CAAAACAATA AATATGATTT TAAATTCTCT TAAAAAAA AAAAAAAACC	1320
45	CCCCCCCCCC GCCCCGN	1337

(2) INFORMATION FOR SEQ ID NO: 186:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

60

GGCACGAGCC TGGACCGAGC AGCCACCGCC CGCTCCCTCT CTCCACGAGG CTGCGGGCTT

60

	AGGACCCCCA GCTCCGACAT GTGCCCTCT GGTCGCCTGT GTCTTCTCAC CATCGTTGGC	120
	CTGATTCTCC CCACCAGAGG ACAGACGTTG AAAGATACCA CGTCCAGTTTC TTCAAGCAGAC	180
5	TCAACTATCA TGGACATTCA GGTCGGACA CGAGCCCCAG ATGCAGTCTA CACAGAACTC	240
	CAGCCCACCT CTCCAACCCC AACCTGGCCT GCTGATGAAA CACCACAACC CCAGACCCAG	300
10	ACCCAGCAAC TGGAGGAAC GGATGGGCCT CTAGTGACAG ATCCAGAGAC ACACAAGAGC	360
	ACCAAAGCAG CTCACTCCAC TGATGACACC ACGACGCTCT CTGAGAGACC ATCCCCAAGC	420
	ACAGACGTCC AGACAGACCC CCAGACCCCTC AAGCCATCTG GTTTTCATGA GGATGACCCC	480
15	TTCCTCTATG ATGAACACAC CCTCCGGAAA CGGGGGCTGT TGGTCGCAGC TGTGCTGTC	540
	ATCACAGGCA TCATCATCCT CACCAAGTGGC AAGTGCAGGC AGCTGTCCCCG GTTATGCCGG	600
20	AATCATTGCA GGTGAGTCCA TCAGAAACAG GAGCTGACAA CCYGCTGGC ACCCGAAGAC	660
	CAAGCCCCCT GCCAGCTCAC CGTGCCCAGC CTCCCTGCATC CCCTCGAAGA GCCTGGCCAG	720
	AGAGGGAAAGA CACAGATGAT GAAGCTGGAG CCAGGGCTGC CGGTCCGAGT CTCCCTACCTC	780
25	CCCCAACCTT GCCCGCCCT GAAGGCTACC TGGCGCCTTG GGGCTGTCC CTCAAGTTAT	840
	CTCCTCTGYT AAGACAAAAA GTAAAGCACT GTGGTCTTTG CAAAAAAA AAAAAAAA	900
30	AAAAAAAAA AAAAAAAA AAAAAAAA AAAAAACTCG A	941

(2) INFORMATION FOR SEQ ID NO: 187:

35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 654 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
40	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:	
45	GAATTCTGGCA CGAGGCAGCT TGTGCTTTAA AGGAGGTGTT CAAAGCATGT CTGAGCAGAG	60
	ACTTTTGGGC TCTGTTTAA TTAATACTTT AAAATAATTG ATATTTAAAA TATCARATGT	120
	TTCCATAAAG AGGAGGAATGT TTAAATGCCT CCAGACTACA TTCCCTTTTA TTSCCTTGATT	180
50	TTACCTGGGA GTCCAAGTT CAATTCCCAT AAAGCAAGCG TTTTATTGT CACTTTCAAT	240
	ATACATCCGA TTGCCATGCT TAAGATGCAA TATGGGCTGC GGAAATAGGT TAACCCACAG	300
	GCTCCCAGGG CCCAGTGTAG AAGGTGAGAG ATTCTGTAA AATGATCAA ATAAAAGGAA	360
55	GACCCCTGGCC GGGTGCCGTA RCTCACGCCT GTAATCCAG CACTTTGGGA GGCGGAAGCG	420
	AGTGGATGAC GAGGTTAGGA GTTGGAGACC AGCCTGGCCA ACATCGTGAA ACCCCGTCTC	480
60	TACTAAAAAT ACAAAAATTA GCCGGGCATG GTGGCAGGCA CCTGTAATCC TAGCTAGTTG	540

GGAGGCTGAG GCAGGAGAAT CGTTGAATC TGGGAGTTGG AGGTTGTCA TGAGCTGAGA 600
 TCGCGCCACA GCACTCCAGC CTGGGTGACA GGGTGAGACT CTGTCTCAA NAGA 654

5

(2) INFORMATION FOR SEQ ID NO: 188:

10

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1848 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

GAAACTGGAC CCGAGAACCG GACCGAAGCG AAGCGGAAGC CCGGAATGAG GCCGGACTGG 60
 AAAGCCGGAG CGGGGCCAGG CGGGCCTCCC CAAAAGCCTG CCCCTTCATC CCACCGGAAA 120
 CCGCCGGGCC GGCGAGCGC GGCGGCCGCT GCGATTGCAG TCGOGGCCGC GGAGGAAGAG 180
 25 AGACGGCTCC GGCAGCGGAA CCGCTGAGG CTGGAGGAGG ACAAAACCGGC CCTGGAGCGG 240
 TGCTTGGAGG AGCTGGCTTT CGCGCACGTC GAGAACGACG AGGACGCGTT GCTGCGGCGT 300
 CTGCGAGGCC CGAGGGTCA AGAACATGAA GACTCGGGTG ACTCAGAACT GGAGAATGAA 360
 30 GCAAAAGGTA ATTTTCACC TCAAAAGAAG CCAGTTGGG TGGATGAAGA AGATGAAGAT 420
 GAGGAAATGG TTGACATGAT GAACAATCGG TTTCGGAAGG ATATGATGAA AAATGCTAGT 480
 35 GAAAGTAAAC TTTCGAAAGA CAACCTTAAA AAGAGACTTA AAGAAGAATT CCAACATGCC 540
 ATGGGAGGAG TACCTGCCTG GGCAGAGACT ACTAAGCGGA AAACATCTTC AGATGATGAA 600
 40 AGTGAAGAGG ATGAAGATGA TTTGTTGCAA AGGACTGGGA ATTTCATATTC CACATCACT 660
 TCTCTTCCAA GAGGCATCTT GAAGATGAAG AACTGCCAGC ATGCGAATGC TGAACGTCT 720
 ACTGTTGCTC GGATCTCCAT CTGTGCAGTT CCATCCCCGT GCACAGATTG TGATGGTTGC 780
 45 TGGGATTAGA TAATGCTGTA TCACTATTC AGGTGATGG GAAAACAAAT CCTAAATTC 840
 AGAGCATCTA TTGGAAAGG TTTCCAATCT TTAAGGCTTG TTTTAGTGCT AATGGGGAA 900
 50 AAGTTTACG CACGAGTACC CACAGCAAGG TTCTTTATGT CTATGACATG CTGGCTGGAA 960
 AGTTAATTCC TGTGCATCAA GTGAGAGGTT TGAAAGAGAA GATAGTGAGG ACCTTTGAAG 1020
 TCTCCCCAGA TGGGTCTTC TTGCTCATAA ATGGCAATGC TGGATATTG CATTGCTAG 1080
 55 CAATGAAGAC CAAAGAACTG ATTGGAAGCA TGAAAATAA TGGAAGGGTT GCAGCATCCA 1140
 CATTCTCTTC AGATACTAAG AAAGTATACG CCTCTTCGGG GGATGGAGAA GTTTATGTT 1200
 60 GGGATGTGAA CTCAGGAAG TGCCTTAACA GATTGTTGA TGAAGGCAGT TTATATGGAT 1260

441

	TAAGCATTGC CACATCTAGG AATGGACAGT ATGTTGCTTG TGGTTCTAAT TGTGGAGTGG	1320
	TAAATATATA CAATCAAGAT TCCTGTCTCC AAGAAACAAA CCCAAAGCCA ATAAAAGCTA	1380
5	TAATGAACCT GGTTACAGGT GTTACTTCCTC TGACCTTCAA TCCTACTACA GAAATCTTGG	1440
	CAATTGCTTC AGAAAAAAATG AAAGAAGCAG TCAGATTGGT TCATCTTCCT TCCTGTACAG	1500
10	TATTTCAAA CTTCCCAGTC ATTAAAAATA AGAATATTTTC TCATGTTCAT ACCATGGATT	1560
	TTTCTCCGAG AAGTGGATAC TTTCCTTGG GGAATGAAAA GGGCAAGGCC CTGATGTATA	1620
	GGTGCACCA TTACTCAGAC TTCTAAAGAG ACTATTTGAA GTCCAGTGA GTCACAAGAG	1680
15	AAGCCTGTCT TGATATATCA TCTCAGAAC TTTCTGAAT ATGTGATAAT ATATGGAAAA	1740
	TGATTTATAG ATCCAGCTGT GCTTAAGAGC CAGTAATGTC TTAATAAACCA TGTGGCAGCT	1800
20	TTTGTGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAACTCGA	1848

20

25 (2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1146 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

	AAAAAAAACC CAGGGGAACN TTGGGGGCCG CTTTNNNTTC CCCCTCCAGG CCATTGGGA	60
35	ATTCTTCAAG TTAATCCTGC TTTCCTTGG GCCAACAGGG CTTGTAGGGG GGAGAGACCC	120
	AGGATCATCA AGGGGTTCGA GTGCAAGCCT CACTCCCAGC CCTGGCAGCC AGCCCTGTTC	180
40	GAGAAGACGC GGCTACTCTG TGGGGCCACG CTCACTGCC CCAGATGGCT CCTGACACCA	240
	GCCCACIGCC TCAAGCCCCG CTACATAGTT CACCTGGGC ACCACAACCT CCAGAAGGAG	300
45	GAGGGCTGTG AGCAGACCCCG GACAGCCACT GAGTCCTTCC CCCACCCCGG CTTCAACAAAC	360
	AGCCTCCCCA ACAAAAGACCA CCCCAATGAC ATCATGCTGG TGAAGATGGC ATGCCAGTC	420
	TCCATCACCT GGGCTGTGCG ACCCCTCAC CTCTCTCAC GCTGTGTAC TGCTGGCACC	480
50	AGCTGYCTCA TTTCCGGCTG GGGCAGMACG TCCAGCCCC AGTTACGCCT GCCTCACACC	540
	TTGSGATGCG CCAACATCAC CATCATTGAG CACCAGAAGT GTGAGAACGC CTACCCCGC	600
	AACATCACAG ACACCATGGT GTGTGCCAGC GTGCAGGAAG GGGCAAGGA CTCCCTGCAG	660
55	GGTGACTCCG GGGCCCTCT GGTCTGTAAAC CAGTCTCTTC AAGGCATTAT CTCCCTGGGC	720
	CAGGATCCGT GTGCGATCAC CCGAAAGCCT GGTGTCTACA CGAAAGTCTG CAAATATGTG	780
60	GACTGGATCC AGGAGACGAT GAAGAACAAAT TAGACTGGAC CCACCCACCA CAGCCCATCA	840

CCCTCCATTT	CCACTGGTG	TTTGGTCCT	GTCACCTCG	TTAATAAGAA	ACCCTAACGCC	900	
AAGACCCCT	ACGAACATTC	TTTGGGCCTC	CTGGACTACA	GGAGATGCTG	TCACTTAATA	960	
5	ATCAACCTGG	GGTCGAAAT	CAGTGAGACC	TGGATTCAA	TTCTGCCTTG	AAATATTGTG	1020
	ACTCTGGAA	TGACAACACC	TGGTTTGTTC	TCTGTTGTAT	CCCCAGCCCC	AAAGACAGCT	1080
10	CCTGGCCATA	TATCAAGGTT	TCAATAAATA	TTTGCTAAAT	GAAAAARAAA	AAAAAAAAAA	1140
	ACTCGA						1146

15

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:

20	(A) LENGTH: 906 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

ACTCCCTCAC	CCAGGTCCCCA	GCCCTGGAA	CCACCTACCG	TGAGCCCTTT	TGCAGATATA	60	
30	GACTCATTTTC	ATCCTCAGAT	GGTCCTCAA	GGTAGGTACT	TTAGTCCCCT	TTAGAGATG	120
	AGACGATTGA	GGCCAGAGGG	GTCAGNTAAC	TTGCCTGGGG	GTCACGAGC	ACAAAAGGAG	180
	CCGAGGCAGG	ATCTGACCCCT	TGTTCTCTGG	CCTCACTGCC	CTCACTTTGC	CATGACCCGA	240
35	AGTTATGTCC	CTACAAAGCA	ATGCATGGTC	CAAGGYCTTT	TTTATTGTAT	TTTTATTTTT	300
	AAGGGTCCTG	TTCAAAACTG	GTGTGAGCTC	TGAGGAGTCC	TGAACCCCTGG	GTGCAGGCATC	360
40	CTAGCATCCT	GGGAGTCCTT	TTCTGCCAC	ACTGAGCTGG	GCTCCTCGAG	GGGTGGGGCT	420
	GCTGTCCCTG	GAAGCCTGGC	AGCAGCACTG	TATCGGGTTG	GCTGAAGCTG	ARCGCCGTGG	480
	GGTGCAGGGC	TCCMGAATC	CCCCTTGGC	TGAAGGGTT	CCCTGTAGCC	MGGGATGTTT	540
45	ATGAGGTCTC	TCTGATGCC	CAGGCGCAGG	ACATGTGTGC	GGGTGGAGAA	AAGCAGGCC	600
	TTTCAGTGCC	AGCTCCACTC	AATTTCTATG	TGGACCAAGA	ACGATAAACT	AAAAAAATT	660
50	TTTTCTCAA	GGTATCTCA	GAATATGGTG	TATTTTATG	TGGAAAAGAA	AAGTTATGAA	720
	GGCAGCTGTT	ACTTTAAGAG	AAAATTCACT	AAAAGTCCTC	GAGGTATGAA	GATGACGGCG	780
	TGCTTCTCAA	TCATTTGGC	ATAACTTGAT	TGTGGCTGTA	ATTTTTTTT	TTTTTTTGT	840
55	CAAGCATGTC	AGACAATAAA	GTCTTGTAA	AAAGRGA	AAAAAA	AAAAAA	900
	ACTCGA						906

60

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xii) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

	CTTCAGCTGA AGCCCAGGGA CCCCTTTTCC ACCCTGGGCC CCAATGCCGT CCTTTCCCCG	60
15	CAGAGACTGG TCTTGAAAC CCTCAGCAAA CTCAGCATCC AGGACAACAA TGTGGACCTG	120
	ATTCTGGCCA CACCCCCCTT CAGCCGCCTG GAGAAGTTGT ATAGCACTAT GGTGCGCTTC	180
	CTCAGTGACC GAAAGAACCC GGTGTGCCGG AGATGGCTGT GGTACTGCTG GCCAACCTGG	240
20	CTCAGGGGGA CAGCCTGGCA GCTCGTGCCA TTGCAGTGCA GAAGGGCAGT ATCGGCAACC	300
	TCCCTGGCCTT CCTAGAGGAC ACCCTTGCGG CCACACAGTT CCAGCAGAGC CAGGCCAGCC	360
25	TCCTCCACAT GCAGAACCCA CCCTTGAGC CAAYTAGTGT GGACATGATG CGGCGGGCTG	420
	CCCGCGCGCT GCTTGCCTTG GCCAAGGTGG ACGAGAACCA CTAGAGTTT ACTCTGTACG	480
	AATCAOGGCT GTTGGACATC TCGGTATCAC CGTTGATGAA CTCAKTGGTT TCACAAGTCA	540
30	TTTGTGATGT ACTGTTTTTG NATTGGCCAG TCATGACAGC CGTGGGACAC CTCCCCCCCC	600
	CGTGTGTGTG TGGGTGTGTG GAGAACTTAG AACTGACTG TTGCCCTTTA TTTATGCAA	660
35	ACCACCTCAG AATCCAGTTT ACCCTGTGCT GTCCAGCTTC TCCCTGGGA AAAAGTCTCT	720
	CCCTGTTCTC TCTCTCCTT CCACCTCCCC TCCCTCCATC ACCTCACGCC TTTCTGTTC	780
	TTGTCCTCAC CTTACTCCCC TCAGGACCCCT ACCCCACCCCT CTTGAAAAG ACAAAAGCTCT	840
40	GCCTACATAG AAGACTTTTT TTATTTAAC CAAAGTTACT GTTGTGTTACA GTGAGTTGG	900
	GGAAAAAAA TAAAATAAAA ATGGCTTTCC CAGTCCTTGC ATCAACGGGA TGCCACATT	960
	CATAACTGTT TTTAATGGTA AAAAAAAA AAAAAATAC AAAAAAAAT TCTGAAGGAC	1020
45	AAAAAAGGTG ACTGCTGAAC TGTGTGTGGT TTATTGTTGT ACATTCACAA TCTTGAGGAA	1080
	GCCAAGAAGT TCGCAGTTGT GAACAGACCC TGTTCACTGG AGAGGCCTGT GCAGTAGAGT	1140
50	GTAGACCCCT TCATGTACTG TACTGTACAC CTGATACTGT AAACATACTG TAATAATAAT	1200
	GTCTCACATG GAAACAGAAA ACCCTGGTC AGCAGCAAGC TGTAGTTTTT AAAAAATGTTT	1260
55	TTAGTTAAC CTTGAGGAGA AAAAAAAA AGGCTTTCC CCCAAAGTAT CATGTGTGAA	1320
	CCTACAAACAC CCTGACCTCT TCTCTCCTC CTTGATTGTA TGAATAACCC TGAGATCACC	1380
	TCTTAGAACT GGTTTTAACCC TTTAGCTGCA GCGNCTACGT CNAWCNTGT GTATATATAT	1440
60	GACGTTKGTAC ATTGCACATA CCCTTGGATC CCCACAGMTK GGTCCTCCTC CCAGCTACCC	1500

	CTTTATAGTA TGACCGAGTTA ACAAGTTGGT GACCTGCACA AAGCGAGACA CAGCTATTAA	1560
	ATCTCTTGCC CAGATATCGC CCCCTCTGGT GCGATGCTGT ACAGGTCTCT GTAAAAAGTC	1620
5	CTTGCTGTCT CAGCAGCCAA TCAACTTATA GTTTATTTTT TTCTGGGTTT TTGTTTGT	1680
	TGTTTTCCT TCTAATCGAG GTGTGAAAAA GTTCTAGGTT CAGTGAAGT TCTGATGAAG	1740
10	AAACACAATT GAGATTTTTT CAGTGATAAA ATCTGCATAT TTGTATTCAC ACAAATGTAGC	1800
	TAAAACCTGAA TGTAATTCCT CCCTTTTTT CCTTTTTGTT CTAAATGAAT ATCAATTATT	1860
	CAGTATGAAA TCTTTATACT ATATGTTCCA CGTGTAAAGA ATAAATGTAC ATTAAATCTT	1920
15	GGTAAGACTT TAAAAAAAAA A	1941

20

(2) INFORMATION FOR SEQ ID NO: 192:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2118 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:	
30	AAATAATAAT AANAATAAAT AAAAATWAAG TGCTTAKTGT AACTCAGCGG ACAGGGCTCC	60
	CAGCTGCTCT GGACACGTGGG ACACCYTCCA CCCTGCACAC AACAGGCATG CAAAGAGGAC	120
35	TGGATATGGT GGGGTAGAGT GCTTCTGGTG TGTTCACTTT AAGAAAACAT CTGCCAAGAG	180
	AGAAGAGTGC CCAGGAAAGA CCAGGAAAT ACAAGTACAT GGCTGCTTCA TACCATATAC	240
40	CCCAATTCTT TAAAGCAGCA AAAGGCACCTT TTTTTTCAG GCCAGAGTGA ATCTAAAACA	300
	AACCTGGCTT TGCTTACAGG GAAGCTGTCC CAGAAGGACT GAGTGATGCC TCTTGTCCC	360
	TAAGGTCTGG AGAGCTTTTG CAAGTTCCA ACGACATTTC CAACCAGGTG GGAGAGACCA	420
45	GCAGTTGACG AGACAAGTCA GACCCAAAAA ACGACGCCAA GGTAGTGAGT GGGTGCCTAT	480
	TTGGGAGTAG GATGATTGAA GGAAAACAGG AAGAAAAACC GGTCAGAAAG TGGCACTTIG	540
	GAAGTGGAAA GCTGTTGCA AATAGCAACT CTGGCTAAAG CGAAAATGTT AATCAAGTAG	600
50	AAAGTAAAAT TCAGGATCTT AGAAGCTCAT CCTTCTGATG AGAACTATTT TTTTTTCCGT	660
	GAAGGAACTA TTATTACTTT AAAAGTGAGG GTAATTACA TATGGGGTGT ATATATTCTA	720
55	AAAATAGTAA TAAAAGTACC TTTTATAAGC AATGTTGTGT GGCTTGTAGA AGAAAGCAGG	780
	GAGGAAAAAA AGGCAGGCCAA AACTAGTCTA GGTCTAGGCC CTAAAAATGAA GCTTCCCTCC	840
60	CACTTGACTG GAAACGCCCA TGTGATTCT AGGCTGAAAA TAGGTAGGAT TTAACGAGTA	900

	ACCTAGTTCC CTCTGTC TCATTTCTGA TCAGCTGATG GAGCTGCTAG TAAGAGGGC	960
	CGATCATGCT CCCAGACGAG TCCCTTGGCC TCTTGCTCTC CATCCCAAGC CTGACTCCTT	1020
5	CAGCAGCAGC CCCCTCCCTC TGTCAGGCATC TGATGCAGGC AAGCAGGAGC AGTAAGAGGG	1080
	CATCCCATGT TCCAGTCAC CTTCTATGGG GTGACTARGA GGTTCCCCGT AACTAGGGCA	1140
10	GCCCARGCCC AGCAGGTGCA AAAAGCAGCT GCAAGCTTCA GAAACCCACT TCCTCCAACA	1200
	CCAGGGAGGT GGCAGAGAGC CCATCCAAAA GCCCACTGGG AGAGGCATAA GATTCTGTGC	1260
	CAGGCCCTCA GGTCCTCTCT GTGTCAGGTA GGCTCTGCTA CTGGCTCTG AAGTAAAGGC	1320
15	AAANACAAAC GGGCAGGGCA GGGTGGCAGG AATAAAAAC TCTGGACAGA AACCCCTTTA	1380
	ATAAAAGAAA TTCCACCCCT CCCAATCCTT CCATGGAAGG GTGAGACCTT AATGTGATGT	1440
	AAGAGGAAGG TCTTCTCTGG CTTTCAGGGA AACAGCTGCA GCTGAAACTT AGGGGCCAT	1500
20	TCCAGGGCAC TTTTCACCAC AGCCAGTGCA GCCGCTCAA GTGCCACTGT CAGCCCCATC	1560
	ACTGCCAATT TCACAAAGCG GTTGGTCTT GGCTTGGTCA GGACATCTTT TGTTGATCT	1620
25	TCAGGCCGCA GAAGTCCCCG AANACCGCTG CGCGAGCACC ATATCAGGCC TCTGCTGGC	1680
	TGATGCCAGC TCAAAGCTT TGAAAGTAGA GGCTGCCGTC CTCTCAGCTT GCTGTTGGC	1740
	AGCGGCCCTCC CGAGCAAGTT CGGATGGGGG AAACGTAAACA AAAAGGTCTC CTSTCTGCTG	1800
30	ATCAGTGTCT CATAGGGCAA GTCTGAGGG ATCTGGGACA ACAGGTGGTG GACCGAGGCC	1860
	ATGTCACAGT CACAGTCCAG GACTTCCTGC TCGCGATACA ACACAATCAC GGCTGCAAAG	1920
35	TAAATCGGCA TCAGTGGGTG GCAGGCCAGG AAGAAGTCAT ATAACCGCAC GACGTGCCGT	1980
	AAGTCAGACA GGACATGCCA AAACCAGGTG ATGAGCCAGC TGAGGGCAA GATGGTCCCT	2040
	ACCTCAGCAC TCTGCATGAA GTCATGGAGC TCTGGATTCA CCTGGTCAAT GATGGGCATC	2100
40	AGATAGTTTA ATATATGC	2118

45

(2) INFORMATION FOR SEQ ID NO: 193:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1538 base pairs	
50	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:	
55	CGGGGTTTCGG CTCTGTC TCA GCAGCCGGC GGCGCTCGGG CGGGACATGG CAGCCTGTAC	60
	AGCCCGGGGG CCTGGCCGTG GGCAGCCGCT GGTGGTCCCG GTGCGTGAAT GNGGCCCGGT	120
60	GGCCAAGGCC GCTCTGTGCG CGGCCGNAGC TGGAGCCCTTC TCGCCAGCGT CGACCACGAC	180

	GACCGGGAGG CACCTCTCGT CCCGAAACCG ACCAGAGGGC AAAGTGTGG AGACAGTTGG	240
5	TGTTGTTGAG GTGCCAAAC AGAATGGAAA ATATGAGACC GGGCAGCTTT TCCCTCATAG	300
	CATTTTGCG TACCGAGGTG TCGTCCTGTT TCCCTGGCAG GCCAGACTGT RTGACCGGGA	360
	TGTGGCTCT GCAGCTCCAG AAAAACAGA GAACCCCTGCT GCCCATGGCT CCAAGGAGGT	420
10	GAAAGGCAA ACTCACACTT ACTATCAGGT GCTGATTGAT GCTCGTGACT GCCCACATAT	480
	ATCTCAGAGA TCTCAGACAG AAGCTGTGAC CTTCTTGGCT AACCATGATG ACAGTCGGGC	540
	CCTCTATGCC ATCCCAGGCT TGGACTATGT CAGCCATGAA GACATCCTCC CCTACACCTC	600
15	CACTGATCAG GTTCCCCATCC AACATGAAC TTTGAAAGA TTTCTCTGT ATGACCAGAC	660
	AAAAGCACCT CCTTTTGTGG CTCCGGAGAC GCTAAGGGCC TGGCAAGAGA AGAATCACCC	720
20	CTGGCTGGAG CTCTCCGATG TTCATCGGGA AACAACTGAG AACATACGTG TCACITGTCAT	780
	CCCCCTCTAC ATGGGCATGA GGGAAAGCCA GAATTCCCAC GTGTACTGGT GGCGCTACTG	840
	TATCCGTTTG GAGAACCTTG ACAGTGTATGT GGTACAGCTC CGGGAGCGGC ACTGGAGGAT	900
25	ATTCACTCTC TCTGGCACCT TGGAGACAGT GCGAGGCCGA GGGGTAGTGG GCAGGGAAACC	960
	AGTGTATCC AAGGAGCAGC CTGCGTTCCA GTATAGCAGC CACGTCTCGC TGCAGGCTTC	1020
30	CACTGGGCAC ATGTGGGCA CGTTCCGCTT TGAAAGACCT GATGGCTCCC ACTTTGTATGT	1080
	TCGGATTCCCT CCCCTCTCCC TGGAAAGCAA TAAAGATGAG AAGACACCAAC CCTCAGGCCT	1140
	TCACTGGTAG GCCAGCTGAG GCCCCAAGTG CCCAGGCTTG GTCACCGGGGA AGAACAACTC	1200
35	TCATCCCACA ATTGCTGCAG AACCTCTCTC TCCCCATCAT GGGCCACAGT GGGTCTCTTA	1260
	ATTTGATTGT GGGGTCTTT TTGTGGGGAG GGGTGGTATA ACTTTCTTC AGAAGACCCA	1320
40	TGTGGGACAC CTCCAAGGCT GCCCTCTCA TAAGCCCTGC CTACACCATG TTCCAGTAAA	1380
	CCTCTCCACC AAGGAACGTG GTTCAGCTGC CACAGGCCTG GAGGAGTTTC CTGGCTGTC	1440
	ACGTGAGGTT TGATCAGTAA ACCAGTGCAS GTTGGCCAA AAAAAAAA AAAAAAAA	1500
45	AAAAAAAAA AAAAAAAA AAAAAAAA AAACTCGA	1538

50

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1098 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

60

	AGACCCGTGTC TCAAATAATA ATAATAATAA TAATCTTATT TTGGAGAATA AAGAGACCTS	60
	TGGATTGAG GTGCCATTG GGTAGAAAGA AAAGACGTT ACACCGAGAA ATAGTCTGTG	120
5	TTGCCCTGAA GGAGCAGAGG GATGCATCGC TGGAGGTGAC CTACAGTTGA AGAAGACTCA	180
	TTATGACAGA CCTTGTCCTT CTTCTTGTG GAAAGTGTGTT CCTCTGCTGC TACTGCTCAT	240
10	GAGACTCTTC CCCCTCCCTG TCCCAGGGAA CCAAAGGGCT TTNCTACCAC ACCCTTTCTT	300
	NGCCCCCGC CTCCCATGTC TGCTGTGCCT TTGTACTCAG CAATTCTTNG TTGCTCCCA	360
	TTATCTCCA GCCGGATACA GAGTGAATAG TTAACCACAC TTAGGTCAAA TAGGATCTAA	420
15	ATTTTTGTTTC CTGCTCCNGT GTAAAGAGGC CAGTGTGTGTT GTGTGCAAG CAGCCTTGGA	480
	ATAGTAACTC TTCTCATTTG TTTGGGATCT CGCCAMCAAG TTCCAGAATG ATACACGGAT	540
20	CAGTGCAGAA GTTCATCAGG CTCTCGGACC TTAGGGCTGT TGGAGAACCC TTCAGCAGCA	600
	GAACGTATGG TKAWKGVTG TGTTCTCCAT CCTCAACTTT CTTTGCTTCG ATCATAACACA	660
	AGAATACATT TGGAAAGGGCA AAAAATGAAC ACTGTTGTTTC ATTGCAGCCG TGTTTGTGA	720
25	CACAGATGCA CAGTCTGCTG TGAAGACCTT CTCTCAAGTG GSATYTGGGA GTCCATGCCA	780
	GATCATGGTG CTTCATGAGA GACTGACAGC TATCAGGGGT TGTGGCACTT AGTGAGGACT	840
30	CTCCTCCCCC AGTGTGTGCT GATGACACAT ACACACCTGA CAATAGCTTG AGTCTTCTCT	900
	GTTCCTTTTA CTCTGTAGCC AACATACACA TGATTTAAAA CCCTTTCTAA ATATCTATCA	960
	TGGTTCATCC TTGTCCAAT GCAGAGTCAG AGCTATTGTG ACTTCATTAT TATTCCAAG	1020
35	GCGAATAGTT GGCTTCTTT TTGCAAAAAT AATTAAAGTT TTGTATGTT GCAAAAAAAA	1080
	AAAAAAAAAA CTACGTAG	1098

40

(2) INFORMATION FOR SEQ ID NO: 195:

	(i) SEQUENCE CHARACTERISTICS:	
45	(A) LENGTH: 1001 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:	
	GAATTCCGCA CGAGATAGCT TGCATCTCAT CCCAGTAAAA CCACATTATT ATAACATATC	60
	AACGTATTGA CAAGGTGAA GAGCAAGATT GTTCTGAGGT GAGATGCAA TTTCAAAGGG	120
55	GTGAGCACTA ATTGTTCCAG TGATTGTTA TTTATTGGCT AGGACATAAT TACTCTCTT	180
	GAGGTTACAC ATCTGCCTCC AGGTTCTGT GTGCTGTG CTTGGGATC AGGCCAGGGC	240
60	AGACTGTGAT CACTGAGATT CAAACTCCCA GARTAATCAG CAAGAGCTTT CTAGAGACCA	300

	AGGCCAGGCC	TGATCCCTGA	GGGATGCATG	AGAAGGCTTG	GAATTCATT	CTGCTATGGT	360
5	GGCTCTCTCT	TGATCTTCTT	GGAGTAGCAA	AAACAGCAAT	GTGGGCCAA	TGGTGTGGCC	420
	TAAATGATCA	CAAAGGTAAA	TGACTAAAGG	GCTCAGCAGA	TGAGTAAGGA	GCCTTGTCT	480
	GAGAAATTAG	CACTGGGCTC	TGCATTCAGA	AACATGTGAT	AAGCATTGCC	CATTGCACAT	540
10	TGCCTTTATT	GTGTAAGGAC	ATGAAATTCC	AGTTTGCAT	AGCTAGTGAT	GAATACTGA	600
	AGGAAATTGC	AGACATATTT	TATTTTATTT	TTAATTGACA	GATGGAATTG	TATATATTTA	660
	TCATGTACAT	AATCATGCTT	TAAAATATGT	ACATTATGGA	ATGGCTAAAT	CAAACTAACC	720
15	TAGGCATTAT	CTCATATAAT	TGTCACTTTT	GTGGCGAGAA	GACTAAAAAT	CTACCTTTTC	780
	AGCATTTTTA	AAGAATACAA	TGTGTTTAT	TAACAAACAGT	CACCAATTGG	TACACTAGAT	840
20	CTCTTGAACT	TCCTCCCTCTT	ATCTAACTGA	GATCTTGTAA	CCTTTGATAA	CAGCTCCCAA	900
	GCCCTTCCCC	AACCACTGCT	CCACCCGTGG	TAACCACCAT	TCTATTCTCA	ACTTCCTGGT	960
25	AATCACCAATT	CTAGACACAG	GGAAAGACTCT	CTACCCCTCTG	A		1001

(2) INFORMATION FOR SEQ ID NO: 196:

30	(i) SEQUENCE CHARACTERISTICS:						
	(A) LENGTH: 1443 base pairs						
	(B) TYPE: nucleic acid						
	(C) STRANDEDNESS: double						
35	(D) TOPOLOGY: linear						
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:						
40	ATAAACTGAA	ATAGGTCTATG	CAAATATAAA	ATATTATTT	TAAATTATTT	GTCATAAGAA	60
	ACGATGGTGG	CCATATTTTG	CTTTAATAAT	GGAAAAAAATG	TGGTTAGCAT	TCTKTGGAAAG	120
	GTGGTCATCA	GATAGTAGAC	ATTTCTAGG	ATTTATTTCT	ACCTGCATAT	GTGGAAATGT	180
45	GTACTACTTT	AGATTTATWT	AATGGCAGCT	AACTCAGAGG	CATCAAAATG	TGCTAATGGT	240
	GTAATATGGC	CTTTGTCTTG	CTGTYCTGTT	TTGTARGCCT	TCAATCAAGC	ARGGGCAGGG	300
	CCGTACAGTG	AACTTGCTCT	TIGSCAGACG	CCAGCGTCTG	CCCCTGACCC	CGTCTCCACT	360
50	CTCTGTGTCC	TGGAGGAGGA	GCCCCCTGAT	GCYTACCTTG	ATTCACCTTC	TGCGTGCGTT	420
	GTACTGAACT	GGGAAGAGCC	GTGCAATAAC	GGATCTGAAA	TCCCTGCTTA	CACCATTGAT	480
55	CTAGGAGACA	CTAGCATTAC	CGTGGCAAC	ACCACCATGC	ATGTTATGAA	AGATCTCCTT	540
	CCAGAAACCA	CCTACCGGTG	AGTGCAGGG	AGTAGAAATC	TGCATCAGCA	CATCAGCACT	600
60	TGGGGATCTA	AGTAAACCTC	TCGGGGAAAA	TGACCAAGTG	GATGTCATCT	CCCAGCTGTT	660

449

	TCTAAGAGCC CAGATGTCGA GAGTATTGTC TCAACCTTENT CCCTCAGGGCC AGAAGACCG	720
	TGAAAAAGCC ACACTGGTTC AGGGACTCAC CGGACGGTTT TGTGTCCACT TCAACCTGCA	780
5	CCGTCCTCTAC CCCAGAGTGG ACTCARATCC TCAAGTCATC CTCTGAACAT TGTGTGAGA	840
	AATTATAAAA GGGCTTTGGC AATATGTTAG CCCAGGATT TGGCTTCCTC CGAAGATGT	900
	GCCGACNITA ACAGTGGCTT AAATGATGGT AAACCTTTA AGATTTCTA AGCTTGCCA	960
10	TGGAGATAC GTTGACTTTT ATAAACMAC CTATAGTTGT TTAACGAAAT CTAAACAAAT	1020
	ATCTGGAGCT CAGGGGTCA ACTGAGGGAA CACATTTGA GATCAGTTGT TTAAATTA	1080
15	AATGCCAGGT AACCCGTTGA AATTATCAA AACATCTCC ACGTACCGA AGAACCTCA	1140
	GAGGATAGTT CTGTTATGGA GAAGATGAAA TGGTTTASTA GTGTAGGAC TTTGGAAAGG	1200
	TGAGCTTAGA TTTGGATAGT AAAACCTCAA GACCCTTTT AAAAAGTAAAT TTAAGAATGC	1260
20	AGCATAAATA ATTTAATTCA GTGTTAANAT GCGAAAGCTA GTATATTGAG CTCATGTGA	1320
	AAAGAAAATC ACATTGGGAG AATGCCACCT TTTCTTATA AGATAGCTTT CAAGATACCA	1380
25	TTTACAGAG ATGGAAATTG AATAGCTTTA GAAAGCCAA ATGTTTGATC TTGGGGAAA	1440
	AAA	1443

30

(2) INFORMATION FOR SEQ ID NO: 197:

	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 1282 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:	
	GAAAAAAAAA AGTATGACCC AGTAGCTAGG CACCTGGC CGCGCCAAAGT TGACCATAA	60
	AATTAACGT CACAGTATCA TCTTAGAAGT GAAAGAAGCC CCTTTATCCT GGAGTGCCCC	120
45	TCTACCACCA CCTACTGACA AAGAACATGG TGCTTCTGG CATGGGAGAA AGTTCAAGTT	180
	TGCTATGGCT TGTATGTGTC CCCTCAAATT CAAGTGTGCA CAATGTGACA GCATCAAGAG	240
50	GTGGGGCTTT TAAGAGATCA CTAGGCCATG AGGGATCTC TTAGGACTGG GTGAAGGCC	300
	CATAATAAAA GAGGTTTCAG GGAGCATCCT GCTAGCTTGC CTTCTGTATG TGTGAAACACA	360
	GCAAGAAAAGC CCTAGTCAAC AAGTGCCAGC TCTTGTACT TAGACTTCCC AGGCTCCAGA	420
55	ACTGTGAGAA ATACATTCT GTTCCCTTACA AATACCCAG TCTCTGTAT TCTGTTATAG	480
	CAGCACAAAA TGAAGATACC ATACCTGACAC ACCTGACAT TCTTCACAG AGTGTAAATG	540
60	CACTGCTTTA TTCTGGTCTC AGTATTGTGT GCTTAATAG GAAATGAGAA AGGGTGGATC	600

450

AGGCATAGG ATGAACAAGT TACTGCTAGA CCTCTCGCAA TGCCACTAAT GCAAGGATT 660
 GTATTTTCAT CATTNCTTGT CTCTTCGGAA CCTAACACCA TGCTATATA CGCCTTAAT
 5 AGATGTCTAA AACACACCTA AGTATTTGTC TAGAAATTG GTGCAATGTC CAGAAGAAC 780
 CAAAATTCAA AATAATTCA AAGGGCCTAA ACCACTTAA ATGCAATT CTTAGTTT 840
 10 TAATGGTACT ACCACTCTCA AATTAAAT GTCATCTTC GTTCCCTTC CTGGCATTTG 900
 ATTATTTGCT AAAACCTGGT AAACACTTTA ATCCYTTCA ATTCCATTC GACTCTCTT 960
 15 GTCCAGAATT ACTCGCAGAC TAATAGTCAC CTGACTTGT CCCCTGGAT CCGAATTGCT 1020
 GTCTAATTCT GGTTACAAAT AAGTAACCTGC CAAACTAATC TTTCTAAAAA GCAAGCTGA 1080
 TCTCGTCACT CCTTTGCTCA ACAATGTAAA AGCTCCCAAT GTCTCCCAA TAAACACC 1140
 20 TTTCCACTGT GTATACAATA CATCCATGAT CTGTATCCAG CATCAATTG TATTTGCTCA 1200
 CTTTATACAC CACCCCCAT GCCACATCAA ATTAATTAA CCTGAAATTG GCAAGTGCAA 1260
 25 AAAAAAAAAA AAAAAAAACTC GA 1282

(2) INFORMATION FOR SEQ ID NO: 198:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 951 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

40 ATTCGGAAC GAGGACTGAA GTGGGACCGG CGGCAGGGTA GAGGAGGAA GGGGATCTA 60
 TGTGGTAACT AAAGAATGTT TCTGTTTGT TAAATTATGTT GTGTTGTTG TTGTTTGTGTT 120
 TGCTTAAGAG AATCAAAAAC TGAAAAAAAT GAGAATACAG GAAATGGCTC TTGTTTATT 180
 45 TTTTGCTGTG TTTACAGCTT GTAAATGCTC TACTGTCTT GTTCAAGAG AGATTTGTC 240
 ACTGCCAGC TCGTTTGTG TCCTGAGCCC TATGCCAGC CCACCTTAA ATCTTCCT 300
 50 GTTTAGATGT TTGATTTGTT TCTGTTGCT ATTGTTATCT TAAAGGTGAA TAACTCTGAC 360
 ATGCCAGACA TCAAATTAAAG CTCAAATTAA GCTCTCGTT AATGTTAA ACACCTAATT 420
 TATATTCTAA TTGATCCAG CCACATGATGC ATGTAATTCA GCTACTTGT CTAAATAAGC 480
 55 ATATTAATT TCCACATCAG GCCATCAGAT CTTGAGAACC AACAGTTATC TAAAGTCCG 540
 TGTCTACTAA TGTTTACCT GCATGCAGCC TTCATTAAAT TTGTTGCAA ATCAAAAGTG 600
 ATCATTATGT AGTTTCTGGA TTAAAAAAAT TTGTTGTTGA AGTTGCTTTG TAAAGTCAT 660
 60

GTGGAATTAA TGGGACAGTG TGCCCTTTGT GTTAGATGTT AGAGCAAAAG AAAGGGCTTA	720
TAGTGTAGT ATTGGAGCAC TTIGAAGATA GATATTTCA GAAAAGATGT AGGATTTAAA	780
5 AGTTAAATT TAAATTTAG AAAAAGATAT GATGGCAATT GGAAATAGTC ACAATGAAGT	840
TCTTCATCCA GTAGGTGTTT AACAGTGTAA TTTTGCCACT GGTAATGTGT AAACGTGAG	900
TGATTTACAA TAAATGATTA TGAATTCAAA AAAAAAAA AAAAAACTCG A	951

10

(2) INFORMATION FOR SEQ ID NO: 199:

15

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1740 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

TTATTATAAT AATGATGATG ATTCCAAGGA AAAAACCTAC AGCGAATGTT CCATTTCTAC	60
25 CCCGCACGCA GACACTCTCC CTAACACTGA TAACCTGAGC CCCCAGCACT GGACGGAAGA	120
ATGCTGGCGT CTCCGTGTGT ACTGGTCAG GGTCTGCC CCAGCCTTGT CAGGACCCCC	180
30 TGGTGTCCAG AGCCCCCACC CCTCCCGCAA CAAGCAGCTG ATGCCCCAGT GATTCTCTAT	240
ACATTTTCA CCTCGGCCAA TATGTCCAGG AAAACTGCCT ACTTCTCTTT TCTTGCTGG	300
35 AGCCTTCATT GTTCACCCCT ACGGTGAAT ATAGGAATTA ATGCTACAAA ATAAAAGTAA	360
AGCTTACCTG AAAAGTCAT AGTTGGGGC AATGGTATCT ACATCTCCCA CTGTGGGAAA	420
ACCAGCAAAG CATCAAAACT CTCAATTCTC CTGTTACCR A ATGCAGATCT GAATTATAAG	480
40 ATGTTTATGT TTGACCATTG TTTCAACAAT GGGATTITGT TACGAATTAT CCCTTTAACT	540
GAAACCTCA GTTTTACTGT TTACATTATT AGGAAAACAG GGATATCTTT TGAATCTAAA	600
AATTGATGT ACAGCATGTG ATTTTGAAAG TTTACATGTA AAGTCACAGT ATAGGTGAAA	660
45 TAACGTTTGT CATATTTGA GACGTATCCT GCAGCCATGT TTTTACGTGA GTGTTTTAGT	720
CAAAGTACAT GGTAGACAGT CTTTCACAAT AAAAGAAAA GGATTTTTT TCCTCCAAAT	780
50 GTACATTAT CAACCTAATG ATTGATTITTT TTAAAAAGAG ATTCGGCCC AGTCGGTTT	840
ATGAAAGTTC ATTGCCCTAA ACTGTGCTGA TTGTTTTAA TCAAGTTATA AATTTCACAC	900
55 CTAGATCATG TATCTACCAA CTCTCCTGCA TTTTCCAAAA GGCATTGAGC TTAAATATTA	960
GTCTTGCTTA GAGTAGGTAA TCCACTTACA TGCTGCGCTA AAGCCATGCC TTTGAAACTC	1020
CTTGTTTAAA ACATGATATG ATTTTGTGG GCAGTTTCAG AAAAGAAAAC AAACAAACAA	1080
60 AAATCGACCC TTAAATTATT ACTTGCAACT CAACAGATCT CCCTGCGCTA CTGCTTTTC	1140

	CAGGAACTTT ACTTCAGGGC TGTCCAGATT GCAGITGTGC CCCGTGTATG TGGATCTAGT	1200
5	TCACAGAGTC TTTGGAAGCC AGCAGTCGTG CCCCTCGTAT ACTGTCCACT CATTATGT	1260
	AGATTGGTA TCCTCAGCAG CCAGTGTAA CACCACTGTC ACGTAGTTAN CAGATTCATC	1320
	TTTTATGTAT TTAAAGTAAT CCATACTATG ATTTGGTTTT TCCCTGCACC ATTAATTCTG	1380
10	GCATCAGATC AGTTTTTGIG TTGTGAAGTT CTACTGTGGT TTGACCCAAG ACCACAACCA	1440
	TGAGACCCCTG AAGTAAAGAT AAGGTACACA TACATTATTT GAGTAACITGT TTCCTGGGG	1500
15	GCCAACTCTGT GTATGCTTT AGAAGTTTAC AGAATGCTTT TATTTTGTC TATAACAAAC	1560
	AGTCTGTAT TTATTTCTGT TGATAAACCA TTTGGACAGA GTGAGGACGT TTGCCCTGTT	1620
	ATCTCCTAGT GCTAACAAATA CACTCCAGTC ATGAGCCCCG CTTTACAAAT AAAGCACTTT	1680
20	TGATGACTCA MAAAAAAA AAAAAMC YCGGGGGGGG GCCGGTAACC CATTNNCCC	1740

25 (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1707 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

35	GCTTATAGAA GGGAGAGGAG CGAACATGGC AGCGCGTTGG CGGTTTTGGT GTGTCTCTGT	60
	GACCATGGTG GTGGCGCTGC TCATCGTTG OGACGTTCCC TCAGCCTCTG CCCAAAGAAA	120
40	GAAGGAGATG GTGTTATCTG AAAAGGTAG TCAGCTGATG GAATGGACTA ACAAAAGACC	180
	TGTAATAAGA ATGAATGGAG ACAAGTTCCG TCGCCTGTG AAAGCCCCAC CGAGAAATTA	240
	CTCCGTTATC GTCATGTTCA CTGCTCTCCA ACTGCATAGA CAGTGTGTG TTTGCAAGCA	300
45	AGCTGATGAA GAATTCCAGA TCCTGGAAA CTCCGGCGA TACTCCAGTG CATTCCACCA	360
	CAGGATATTG TTTGCCATGG TGGATTTGA TGAAGGCTCT GATGTATTTC AGATGCTAAA	420
50	CATGAATTCA GCTCCAACCT TCATCAACTT TCCTGCAAAA GGGAAACCCA AACGGGGTGA	480
	TACATATGAG TTACAGGTGC GGGGTTTTTC AGCTGAGCAG ATTGCCCGGT GGATCGCCGA	540
	CAGAACTGAT GTCAATATTA GAGTGATTAG ACCCCCCAAAT TATGCTGGTC CCCTTATGTT	600
55	GGGATTGCTT TTGGCTGTTA TTGGTGGACT TGTGTATCTT CGAAGAGTAA TATGGAATTT	660
	CTCTTTAATA AAACGGATG GGCTTTTGCA GCTTTGTGTT TTGTCCTTGC TATGACATCT	720
60	GGTCAAATGT GGAACCATAT AAGAGGACCA CCATATGCC ATAAGAATCC CCACACGGGA	780

CATGTGAATT ATATCCATGG AAGCAGTCAA GCCCAGTTG TAGCTGAAAC ACACATTGTT	840
CTTCTGTTTA ATGGTGGAGT TACCTTAGGA ATGGTGCTTT TATGTGAAGC TGCTACCTCT	900
5 GACATGGATA TTGGAAAGCG AAAGATAATG TGTGTGGCTG GTATTGGACT TGTGTGTTA	960
TTCCTCAGTT GGATGCCTTC TATTTTACA TCTAAATATC ATGGCTACCC ATACAGCTTT	1020
10 CTGATGAGTT AAAAAGGTCC CAGAGATATA TAGACACTGG AGTACTGGAA ATTGAAAAAC	1080
GAAAATCGTG TGCTGTTGAA AAGAAGAATG CAACTTGAT ATTGTGATT ACCCTTTTTT	1140
TTCAAGTGAT TTAAATAGTT AATCATTAA CCAAAGAAGA TGTGTAGTGC CTTAACAAAGC	1200
15 AATCCTCTGT CAAAATCTGA GGTATTTGAA AATAATTATC CTCTAACCT TCTCTTCCCA	1260
GTGAACCTTA TGGAACATTG AATTTAGTAC AATTAAGTAT ATTATAAAAA TTGTAAAAACT	1320
ACTACTTTGT TTTAGTTAGA ACAAAAGCTCA AAAACTACTTT AGTTAACTTG GTCATCTGAT	1380
20 TTTATATTCG CTTATCCAAA GATGGGGAAA GTAAGTCCTG ACCAGGTGTT CCCACATATG	1440
CCTGTTACAG ATAACATACAT TAGGAATTCA TTCTTAGCTT CTTCATCTTT GTGTGGATGT	1500
25 GTATACTTTA CGCATCTTC CTTTGAGTA GAGAAATTAT GTGTGTCATG TGGCTCTTG	1560
AAAATGGAAC ACCATTCTTC AGAGCACACG TCTAGCCCTC AGCAAGACAG TTGTTCTCC	1620
30 TCCTCCTTGC ATATTCCTA CTGAAATACA GTGCTGTCTA TGATGTTTT TGTTTGTG	1680
TTTTTYGAG ATCACGYTAC TGGGCTC	1707

35 (2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 779 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

45 CTGTCCCCAG TGTTCCAGG TAATGACTTG GCACICCAGA GAAAGTTTCA TRCTGTGCG	60
TGTGGTGGCT CCAAGCCAAAG CACCTGGCAT GCAGGTTCAGC CCTTCCCAGC GGGCGTGGCG	120
50 TCGTCCTCTT CACAGATGCC ACGTTGCAGC CCCAAGGCCT CACCATTITG CGTTTTTTAG	180
AAACCCATT TCTTGGTCAT TTATAAAGCT GCTTTATAGA TATCTTTGAT CCTGGCATGC	240
55 CTTGGTTTCC TCTCCCTTCC CTTCTTCCAA TCCTGGTTTC CTAACCTCCT CTTGTAGTAA	300
TCTCTCAACTC AACTCAAAGT CCCAAGAATT TGGAAATGGTA GGATGCTGTG CGGGGAGCTC	360
GAGGCTGAGG CATAATCACT GCTTCGGTTC TGTCATCAG GGGACACGCT CCCTTACTCA	420
60 TGGCAGCCAT GTTTGATTGT CACAGAGCCC CCCGAATACT CTGCTATAG TGACACACTG	480

TAGGTGTCAT AAATTTAAG AAACCTGCTT TTAAGTACTA TTTATAGGTT TTCTGTTAT	540
ACTTGCAACC TAGTTTAAA ATACATGAGG ATTTATGAA AGCTTATAC AGACATTTAT	600
5 AGGAAACTCA TTCTTGTATT TTAGGTGCCA TTTAAATTGA TAACACTTAC TTTATAAAAA	660
GATGCTTTT GTCTGGATAG AGCCTTATAG TTTAAATAT CTTCATATAT TGCCATTGTA	720
10 TCAAATAAAT TTCTTACTTA GAAAAAAAAA AAAAAAAAAA AAAACTCGA	779

15 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1617 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

25 GGCACAGCTT TCTGTCCTT CCTCGCTCCC TCTCTTCTC TCCCTCCCT GCCTTCCCAG	60
TGCATAAAAGT CTCTGTCGCT CCCGGAACTT GTGGCAATG CCTATTTTTT GGCTTCCCC	120
30 CGCGTTCTCT AAACTAACTA TTTAAAGGTC TCGGGTCGCA AATGGTTGTA CTAAACGTAG	180
GATGGGACTT AAGTGAACG GCAGATATAT TTCACTGATC CTCGCGGTGC AAATAGCGTA	240
TCTGGTGCAG GCCGTGAGAG CAGCGGGCAA GTGCGATGCG GTCTCAAGG GCTTTTCGGA	300
35 CTGTTTGTCA AAGCTGGCG ACACATGGCC AACTACCCGC AGCCTGGGAC GACAAGACGA	360
ACATCAAGAC CGTGTGACA TACTGGGAGG ATTTCACAG CTGCACGGTC ACAGCCCTTA	420
40 CGGATTGCCA GGAAGGGCG AAAGATATGT GGGATAAACT GAGAAAAGAA TCCAAAAACC	480
TCAACATCCA AGGCAGCTTA TTGAACTCT GCAGCGGG CAACGGGGCG GCGGGGTCCC	540
TGCTCCCGGC GTTCCCGGTG CTCTGGTGT CTCTCTCGGC AGCTTTAGCG ACCTGGCTTT	600
45 CCTCTGAGC GTGGGGCCAG CTCCCCCGC GCGCCCACCC AACTCACTC CATGCTCCCG	660
GAAATCGAGA GGAAGATCCA TTAGTTCTTT GGGGACGGTG TGATTCTCTG TGATGCTGAA	720
AACACTCATA TAGGATGTG GAAAATCCTG ATTCTCTTTT TTATTTCGTT TGATTTCTTG	780
50 TGTTTTATTT GCCAAATGTT ACCAACAGT GAGCAAGCAA GCACAGCCAA AATCGGACCT	840
CAGCTTTAGT CCGCTTCAC ACACAAATAA GAAAACGGCA AACCCACCCC ATTTTTAAT	900
55 TTTATTATTA TTAATTTTTT TTGTTGGCAA AAGAATCTCA GGAACGGCCC TGGGCACCTA	960
CTATATTAAT CATGCTAGTA ACATGAAAAA TGATGGCTC CTCTTAATAG GAAGGCGAGG	1020
60 AGAGGAGAAG GCCAGGGAA TGAATTCAAG AGAGATGTCC ACGGACGAAA CATAACGGTGA	1080

ATAATTCAAG CTCACGTCGT TCTTCCACAG TATCTTGTT TGATCATTTC CACTGCACAT	1140
TTCTCCTCAA GAAAAGCGAA AGGACAGACT GTTGGCTTTG TGTTGGAGG ATAGGAGGGA	1200
5 GAGAGGGAAG GGGCTGAGGA AATCTCTGGG GTAAGAGTAA AGGCTTCCAG AAGACATGCT	1260
GCTATGGTCA CTGAGGGGTT AGCTTATCT GCTGTTGTTG ATGCATCCGT CCAAGTTCAC	1320
TGCCCTTATT TTCCCTCCCTC CCTCTTGTT TAGCTGTTAC ACACACAGTA ATACCTGAAT	1380
10 ATCCAACGGT ATAGATCACA AGGGGGGAT GTTAAATGTT AATCTAAAAT ATAGCTAAAA	1440
AAAGATTTG ACATAAAAGA GCCTTGATTT TAAAAAAAAA AGAGAGAGAG ATGTAATTAA	1500
15 AAAAGTTTAT TATAAATTAA ATTCAAGAAA AAAAGATTTG CTACAAAGTA TAGAGAAGTA	1560
TAAAATAAAAA GTTATTGTTT GAAAAAAA AAAAAAAW CTCGACCGCA AGGAAAT	1617

20

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1974 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GAATTGGCA CGAGGCTGAG GGAGCTGCAG CGCAGCAGAG TATCTGACGG CGCCAGGTTG	60
CGTAGGTGCG GCACGAGGAG TTTTCCCGGC ACCGAGGAGG TCCTGAGCAG CATGGCCCCG	120
35 AGGAGGCCTT TCCCTGGCGC CGCGCTCTGG CTCTGGAGCA TCCTCCGTG CCTGCTGGCA	180
CTGGGGCGG AGGCCGGGCC GCCGCAGGAG GAGAGCCTGT ACCTATGGAT CGATGCTCAC	240
40 CAGGCAAGAG TACTCATAGG ATTTGAAGAA GATATCTGA TTGTTTCAGA GGGGAAATG	300
GCACCTTTA CACATGATTT CAGAAAAGCC CAACAGAGAA TGCCAGCTAT TCCTGTCAAT	360
45 ATCCATTCCA TGAATTTCAC CTGGCAAGCT GCAGGGCAGG CAGAATACTT CTATGAATT	420
CTGTCCCTGC GCTCCCTGGA TAAAGGCATC ATGGCAGATC CAACCGTCAA TGTCCCTCTG	480
CTGGGAACAG TGCCACACAA GGCACTCAGTT GTTCAAGTTG GTTCCCAG TCTTGGAAAA	540
50 CAGGATGGG TGGCAGCATT TGAAGTGGAT GTGATTGTTA TGAAATTCTGA AGGCAACACC	600
ATTCTCCAAA CACCTCAAAA TGCTATCTTC TTTAAAACAT GTCAACAAGC TGAGTGCCCA	660
55 GGCAGGGTGCC GAAATGGAGG CTTTTGTAAT GAAAGACGCA TCTGCGAGTG TCCTGATGGG	720
TTCCACGGAC CTCACTGTGA GAAAGCCCTT TGTACCCAC GATGTATGAA TGGTGGACTT	780
TGTGTGACTC CTGGTTCTG CATCTGCCA CCTGGATTCT ATGGAGTGAA CTGTGACAAA	840
60 GCAAACGTGCT CAACCACCTG CTTTAATGGA GGGACCTGTT TCTACCCCTGG AAAATGTATT	900

	TSCCCTCCAG GACTAGACGG AGAGCA GTGT GAAATCAGCA AATGCCACCA ACCCTGTCGA	960
5	AATGGAGGTAA ATGCATTGG TAAAAGCAAA TGTAAGT KTT CCAAAGGTTA CCAGGGAGAC	1020
	CTCTGTTCAA AGCCTGTC TG CGAGCCTGGC TGTGGTGCAC ATGGAACCTG CCATGAACCC	1080
	AACAAATGCC AATGTCAAGA AGGTTGGCAT GGAAGACACT GCAATAAAAG GTACGAAGCC	1140
10	AGCCTCATAAC ATGCCCTGAG GCCAGCAGGC GCCCAGCTCA GCCAGCACAC GCCTTCAC TT	1200
	AAAAAGGCCG AGGAGCGGCG GGATCCACCT GAATCCAATT ACATCTGGTG AACTCCGACA	1260
15	TCTGAAACGT TTAAAGTAC ACCAAGTTCA TAGCCTTGT TAACCTTCA TGTGTTGAAT	1320
	GTTCAAATAA TGTTCAT TAC ACTTAAGAAT ACTGGCCTGA ATT TATTAG CTTCA TTATA	1380
	AATCACTGAG CTGATATTTA CTCTTCCTTT TAAGTTTCT AAGTACGTCT GTACCATGAT	1440
20	.GGTATAGATT TTCTTGTTC AGTGCTTGG GACAGATTTT ATATTATGTC AATTGATCAG	1500
	GTTAAAATTT TCAGTGTGTA GTGGCAGAT ATT TTCA AAAA TTACAATGCA TTATGGGT	1560
25	CTGGGGCAG GGGAACATCA GAAAGGTTAA ATTGGGAAA AATGCGTAAG TCACAAGAAT	1620
	TTGGATGGTG CAGTTAATGT TGAAGTTACA GCATTTCAGA TTTTATTGTC AGATATTAG	1680
	ATGTTTGTAA CATT TTTAAA AATGCTCTT AATT TTTAAA CTCTCAATAC AATATATT	1740
30	GACCTTACCA TTATTCCAGA GATTCA GTAT TAA AAAAAAAA AAAATTACAC TGTGGTAGTG	1800
	GCATTAAAC AATATAATAT ATTCTAAACA CAATGAAATA GGGAAATATAA TGTATGAACT	1860
35	TTTTGCATTG GCTTGAAGCA ATATAATATA TTGTAAACAA AACACAGCTC TTACCTAATA	1920
	AACATT TTTAT ACTGTTGTAA TGTATAAAAT AAAGGTCGTG CTTAGTTT CTGA	1974

40

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1057 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

50	CGGCCTTCCG GGGCAACCGT TCGTCCCAC NCGGGAAAGG GTCCCTGGAGN CGGGAACTAG	60
	GAGCCTCGGA AGTCCAAGGG CGGAGCGCCC TTTGCTAATA AGCCAACTCAG AACGTGAGAC	120
55	GCTCCGGTGG GNCGGTGCCG TCGAGCGCGG GGTGGAGTCT CGGTGACTTG GCTGGCGGGA	180
	TCAAGTGCAG CTGCTTCAGG CTGAGGTGGC AGATAGTGTAG CGCTGGTGGC GGAGTTAAAG	240
60	TYAAAGCAGG AGAGTAATWA TGAATAGCGC AGCGGGATTG TCACACCTAG ACCGTGCCGA	300

	GCGGGTTCTC AAGTTAGGGG AGAGTTTCGA GAAGCAGCCG CGCTGCGCTT CCACACTGTG	360
	CGCTATGACT TCAAACCTGC TTCTATTGAC ACTTCTTCTG AAGGATACCT TGAGKTTGGC	420
5	GAAGKTGAAC AGKTGACCAT WACTCTGCCM AATATAGAAA GTTGAAGGAA GCAGTAAAAT	480
	TCAGTATCGT AAAGAACAAAC AGCAACAACA ATGTGGAATT CASCCAGGAC TCCCAATCTT	540
10	GTAAAAACATT CTCCATCTGA AGATAAGATG TCCCCAGCAT CTCCAATAGA TGATATCGAA	600
	AGAGAACTGA AGGCAGAAC TAGTCTAATG GACCAGATGA GTAGTTGTGA TAGTTCATCA	660
	GATTCCAAAAA GTTCATCATC TTCAAGTAGT GAGGATAGTT CTAGTGACTC AGAAGATGAA	720
15	GATTGCAAAT CCTCTACTTC TGATACAGGG NAATTGIGTC TCAGGACATC CTACCATGAC	780
	ACAGTACAGG ATTCCCTGATA TAGATGCCAG TCATAATAGA TTTGAGACA ACAGTGGCCT	840
	TCTGATGAAT ACTTTAAGAA ATGATTGCA GCTGAGTGA TCAGGAAGTG ACAGTGATGA	900
20	CTGAAGAAAT ATTTAGCTAT AAATAAAAAT TTATACAGCA TGTATAATTT ATTTTGTATT	960
	AACAATAAAA ATTCCCTAAGA CTGAGGGAAA TATGTCTAA CTTTGATGA TAAAAGAAAT	1020
25	TAAATTGAT TCAGAAAAAA AAAAAAAAAA AACTCGA	1057

30 (2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

40	GAATTGGCA CGAGTCATCC CTCTCCCTCT TTCACTCCCT TACTCTTACT CTGTTTTTG	60
	TGCTCCAGAC AGACAGACCC TACCTCTTTT GCTTCTTTTG TGTTTGTGTTG TTTTGAGATG	120
	GAGTGTGCGCT CTTGTTGCCG AGGCTGGAGT GCAGTGGCGC AATCTGGCT CACCACAAACC	180
45	TCTGCCTCCC GGGTCAAGC AATTCTCCCTG CCTCACCCCTC CCGAGAAGCT GGGGATTACA	240
	GGCATGCGCC ACCACACCCA GCTNAATTTT ATATTTTAG TAGAGATGGT GTTTCTCCAT	300
50	GTTGGTCAGG CTGGCTCAA ACTCCCCAACC TCAGGTGATN CGGCCCTGCTT TGGCCTCCCC	360
	AAAGTGTGG GATTACAGGC GTGAGCCACT GCGCCCGAGCC TCTTTGCTC CTTTATACTC	420
	ATTAACTCAC GCCTGTAATC CCTGTTTG GAGGCCAAAG TGAGAAGGTT GCTTGAGGCC	480
55	AAGAGTTTGA GACTAGCCTG CGCAACACAG CAAGATGCCA TCTTTATAAT AAAAATAAAA	540
	ATAAAAATCA ATTAGCTGGG CATGGTGGAA CGCACCTGTA GTCCCGAGCCA ATTGAGAGGC	600
60	TGAAGTGGGA GGATCATTGA GCCCAGGAGT TGAGGTTGCA GTGAGCCATG ATCATGTCAC	660

TACACTCAGC CTGGGCAATA GAGGGACATG TTGTCTCTAA AAAAAAAA AAAAAACTCG 720

5
A

721

(2) INFORMATION FOR SEQ ID NO: 206:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2465 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

CCACCATTAA TCCAACGTGAA GAGGAGTTAC AGGCAGTTCA GAAAATTGTT TCTATTACTG 60

20

AACGTGCTTT AAAACTGTT TCAGACAGTT TGTCTGAACA TGAGAAGAAC AAGAACAAAG 120

AGGGAGATGA TAAGAAAGAG GGACGTAAAG ACAGAGCTTT GAAAGGAGTT TTGCGAGTGG 180

25

GAGTATTGGC AAAAGGATTA CTTCTCCGAG GAGATAGAAA TGTCAACCTT GTTTTGCTGT 240

GCTCAGAGAA ACCTTCAGAA ACATTATTAA GCCGTATTGC AGAAAACCTA CCCAACACGC 300

TTCGTGTTAT AAGCCCTGAG AAGTATGACA TAAAATGTGC TGTATCTGAA GCGGCAATAA 360

30

TTTGAAATTTC ATGTGTGGAA CCCAAAATGC AAGTCACTAT CACACTGACA TCTCCAATTAA 420

TTCGAGAAGA GAACATGAGG GAAGGAGATG TAACCTCGGG TATGGTAAA GACCCACCGG 480

35

ACGTCTTGGG CAGGCAAAAA TGCCTTGACG CTCTGGCTGC TCTACGCCAC GCTAAGTGGT 540

TCCAGGCTAG AGCTAATGGT CTGCAGTCCT GTGTGATTAT CATACCATT CTTCGAGACC 600

TCTGTCAGCG AGTTCCAATC TGGTCTGATT TTCCAAGCTG CGCTATGGAG TTACTAGTAG 660

40

AGAAAGCAAT CAGCAGTGCT TCTAGCCCTC AGAGCCCTGG GGATGCACTG AGAAGAGTTT 720

TTGAATGCAT TTCTTCAGGG ATTATTCTTA AAGGTAGTCC TGGACTTCTG GATCCTTGTG 780

45

AAAAGGATCC CTTTGATACC TTGGCAACAA TGACTGACCA GCAGCGTGAA GACATCACAT 840

CCAGTGCACA GTTTCGATTG AGACTCCTTG CATTCCGCCA GATACACAAA GTTCTAGGCA 900

TGGATCCATT ACCGCAAATG AGCCAACGTT TTAACATCCA CAACAACAGG AAACGAAGAA 960

50

GAGATAGTGA TCGAGTTGAT GGATTTGAAG CTGAGGGAA AAAAGACAAA AAAGATTATG 1020

ATAACTTTTA AAAAGTGTCT GTAAATCTTC AGTGTAAAAA AAACAGATGC CCATTTGTG 1080

55

GCTGTTTTC ATTCTATAATA ATGTCTACAT TGAAAAATT ATCAAGAATT TAAAGGATT 1140

CATGGAAGAA CCAAGTTTT CTATGATATT AAAAATGTA CAGTGTAGG TATTATTGAA 1200

ATGGAAAGAC ACCCAAAAAA AAAAATGTGC TCCGACTAGG GGGAAAACAG TAGTTCCGAT 1260

60

	TTTTCCCAT TATTTTATT TTATTTCTG GTGCCCTAG CTTCCCCCCC TATTTTGIG	1320
	TCTTTTATTA ACTAGTCAT TGCTTATTA AATCTCACT GTATTAATG CAGGATGTG	1380
5	GCTTCAGTTG CTCTGIGTAT TTTGATATTT TAATTTAGAG GTTTGTTTG CTTTTTGACA	1440
	CTAGTTGAA GTTACTTTGT TATAGATGGT ATCCCTTACC CCTTCATAAT ATTTTACAGC	1500
10	AGTACGTTT TTTGTAACGT GAGACTGCAG AGTTTGTGTT TCTATATGTG AAGGATTACA	1560
	ACACAAAAAG TTATCCGCC ATTGGAGTGC TCAGAACTGA ATGTTCTGC AGATCTTGTG	1620
	GCATTTGTCT CTAGTGTGAT ATATAAAGGT GTAATTAAGA CAGAGTTCTG TTAATCTAAT	1680
15	CAAGTTTGTGCT GTTAGTTGTG CATTAGCAGT ATAAAAGCTA ATATATACTA TATGGTCTTG	1740
	CAACAGTTT AAAGCCTCTG CATAATTGAT AATAAAAATG CATGACATTC TIGTTTTAA	1800
20	TAGACTTTA AAATCATAAT TTTAGGTTA ACACGTAGAT CTTTGTACAG TTGACTTTT	1860
	GACATAGCAA GGCCAAAAAT AACTTTCTGA ATATTTTTT CTTGTGTATA AGTGGAAAGG	1920
	GCATTTTCA CATATAAGTG GGCTAACCAA TATTTTCAA AGAACTTCAT CATTGTACAA	1980
25	CTAACACAG TAACTAGCCC TTAATTATGG TGACAGTTCC TTATTGGTGT GTGTGAGATT	2040
	ACTCTAGCAA CTATTACAGT ATAACACAGA TGATCTCTC CACACACCCC ATCACCCAGA	2100
	TAATTTACAG TTCTGTAAC AGTGAGGTTG ATAAAGTATT ACTGATAAAA AATTTATCTAA	2160
30	GGAAAAAAAC AGAAAATTAT TTGGTGTGGC CATCTACCT GCTTATGTCT CCTACACAAA	2220
	GCTAAATATT CTAGCAGTGA TGTAATGAAA AATTACATCT TACTGTGAT ATATGTATGC	2280
35	TCTGGTACAC AGATGTCATT TTGTTGTCAC AGCACTACAG TGAAATACAC AAAAAATGAA	2340
	ATTCAATATAA TGACTTAAT GTATTATATG TTAGAATTGA CAACATAAAC TACTTTGCT	2400
	TTGAAATGAT GTATGCTTCA GTAAAATCAT ATTCAAATTT AAAAAAAAAA AAAAAAAAAA	2460
40	40 CTCGA	2465

45

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1480 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

55	GAATTCGGCA CGAGCTCAAG CTGGCAGGTG GTGGGGGGAG CGGCCGGAGA GGAGCTGCCG	60
	GGAGTTCTG TGCTGCAGGA CATGACACCA GTGGCATATC ACGGCCATGG GGTCTCAGCA	120
60	TTCCGCTGCT GCTCGCCCCCT CCTCCCTGCAG GCGAAAGCAA GAAGATGACA GGGACGGTTT	180

460

	GCTGGCTGAA CGAGAGCAGG AAGAAGCCAT TGCTCAGTTC CCATATGTGG AATTCAACGG	240
5	GAGAGATACC ATCACCTGTC TCACGTGCCA GGGGACAGGC TACATTCAA CAGAGCAAGT	300
	AAATGAGTTG GTGGCTTGA TCCCACACAG TGATCAGAGA TTGCGCCCTC AGCGAACTAA	360
	GCAATATGTC CTCTGTCCA TCTGCTTTC TCTCCTGGCA TCTGGTTGG TGTTTCTT	420
10	CCTGTTCCG CATTCAAGTCC TTGTGGATGA TGACGGCATE AAAGTGGTGA AAGTCACATT	480
	TAATAAGCAA GACTCCCTTG TAATTCTCAC CATCATGGCC ACCCTGAAAA TCAGGAACTC	540
15	CAACTCTAC ACGGTGGCAG TGACCAGCCT GTCCAGCCAG ATTCACTACA TGAACACAGT	600
	GGTGAATTTT ACCGGGAAGG CCGAGATGGG AGGACCGTT TCCTATGTGT ACTTCTTCTG	660
	CACGGTACCT GAGATCCTGG TGCACAACAT AGTGATCTC ATGCGAACTT CAGTGAAGAT	720
20	TTCATACATT GGCTCATGA CCCAGAGCTC CTTGGAGACA CATCACTATG TGGATTGTGG	780
	AGGAAATTCC ACAGCTATTT ACAACTGCT ATTGGTTCTT CCACACAGCG CCTGTAGAAG	840
	AGAGCACAGC ATATGTTCCC AAGGCCTGAG TTCTGGACCT ACCCCCCACGT GGTGTAAGCA	900
25	GAGGAGGAAT TGGTTCACTT AACTCCCAGC AAACATCCTC CTGCCACTTA GGAGGAAACA	960
	CCTCCCTATG GTACCATTAA TGTTCCTCAG AACCAAGCAGA ATCACTGCCT AGCCTGTGCC	1020
30	CAGCAAATAG TTGGCACTCA ATAAAGATTT GCAGAATTAA ATACAGATCT TTTCAGCTGT	1080
	TCTTAGGGCA TTATAAATGG AAATCATAAC GTGGTTCTAG GTTATCAAAC CATGGAGTGA	1140
35	TGTGGAGCTA GGATTGTGAG TGACCTGCAG GCCATTATCA GTGCCCTCATC TGTGCAGAAG	1200
	TCCGAGCAGA GAGGGACCAT CCAAATACCT AAGAGAAAAG AGACCTAGTC AGGATATGAA	1260
	TTTGTTCAG CTGTTCCCAA AGGCCTGGGA GCTTTTGAA AAGAAAGAAA AAAGTGTGTT	1320
40	GGCTTTTTTT TTTTTAGAA AGTTAGAATT GTTTTACCA AGAGTCTATG TGGGGCTTGA	1380
	TTCACCCCTTC ATCCATTGGC TGGAACATGG ATTGGGGATT TGATAGAAAA ATAAACCTTG	1440
	CTTTGATTC AAAAAAAA AAAAAWAAA AAAAACTCGA	1480
45		

(2) INFORMATION FOR SEQ ID NO: 208:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

60

CAGTATTTCCTCCTCAGTACTG TAAGCAAAAG TGGTATGTGTT TTCTTCTTT ATGTCTACTC

60

	TGICCTCTGT GGCCCTCTGG TGTACCCCTC TCTTCCTAGC CATTCACTCT CTCTAGTCAC	120
	CTCCCTAGTA GCTAGTGCTC TCTAAGTTTT TATTTAATTA GAACAACTCC ATTTCCATT	180
5	CAAGGTAGGT CAATGGGGG AAAAGCCTCA TGATTTAAC TGAAGTTAAC AACACAGCTT	240
	TTAAAATGAA AACTCATACT CCAACTCTA AAGTATATT GAGCTGATT GTTCCAAAA	300
10	CAAAGATATG CTGTACCTAA AACTGCTAAA ACAAAAATAT AAAGACAAGG ACTAGGTGAT	360
	TAAGGGAGA GAAAATCAT YICTTTCCA GGAAACCTTT GCTAAAATAA GCAAAACTTG	420
	ANTCTATGCT TCATGGAAAC TGACACAAAG AAAAGAAACT GATGGATTC ACAGGCCTTG	480
15	TTATAGAAAT AGATCTATAA AAAGATCTGT CCACAGGAA TATACACCTT CTCCCTGGTC	540
	TGAACTTCAA TGGGGATTG TCACCTAGGT CTCCATCTAT AGGAATACCT TCACATACCT	600
20	ATCTATTCAAT GCACATATTG TGAAAACAGG TACATACAAA ATTACAACAA AGGAAAAAAA	660
	TTCTATTGAA CACTAAAAA TAGAACAGG CCAGGCACGG TGGCTCATGC TGTAATCCCA	720
	ACAATTGGG AGGCTGAGGC TGGTGGATCA CCTGAGGTCA GGAGTGTGAG ACCAGCTTGG	780
25	CCAACATGGT GAAACCCCGT CACTACTAAA AATACAAAAA AAATTAGCCT GTGTGGTGGC	840
	ACACTCNTAC AATCCNGGCT GACTCGGGAA AN	872

30

•

(2) INFORMATION FOR SEQ ID NO: 209:

	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 1779 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:	
	AATTGCCAAG ACTGCACAAA ATTACAGTGC TAATGTATAT GGTTGCAGTT CACATAAAGA	60
	CAAAAGCATIC TGTTATGAAA TGAGTAGTAA TATTGGGTGG TTGATTGTGTT CTTAGCAGAC	120
45	TTGGCTTCAT WTGGTCTTG AGATAAAATG GCCAGCATAA ATGCTGTTA TATTCACGTT	180
	TTCTCTAGTG TGTGTGTGCA GGCCACAGCA GCATGCCCTT GGTGTAGTCA GTGCCGAAAS	240
50	GGGTCTGTC CTTCTTGAGC CTGCCTGCAG GGATGGTCTC CTTTAAAGC AGGTGTGTG	300
	CAGCATTCAAG TACACTGAAG GTAAGCTAAA CCATCAACAT CTCTGGTGT TTAAGATGTT	360
	ATTTTATTGG AACAACTGAC AAATGAGGGA TGTTAGCTTT GTGGCAGAAT TCCCTGCATG	420
55	TGTGATAACT GATCTTGTGTT TATTTTTGG CATTGCAACT GTGGCATAGT TACAATTCT	480
	GTTTGKTCAT CACATTAAA ATTGGRAGAG AACGOGCTTG AKGGATAGAG CGCCTTCAGK	540
60	GTACTGTTTC TTATTAACCT TACTTTTTTTT AAATCAACTT GCTATAGACT TTATATACAT	600

	TTTGTAAAT ATAGTCCTA GTGACATAGA AACGATGCGT AGTTTCATT TACTAATTAC	660
5	AAATGTTGAG GCCTAATTCT GAAAGTCCTC ATATTTAAAG GCTAGACAAC GTAATGAAAT	720
	TTTTAACTAT TTGTATGTCA TTTGAAAGT GTACTGCTTT ATGGTAAAAG TGTTTTTCAT	780
	TTGTCATTG TTTTCATTAT TTGTGATCAT GTGTCTTC AATACAGGCA TAAACCTTCC	840
10	ACTCTGAAC AAAGCAGCTG CTTTTAAAAA GCGGTAATTG CTTCTTACC TTTTATTCT	900
	TTTGTAAATG AAGCTTTCT TTAAGAATGT GACTTTAAAG TGTTGTCTAT TGCATAAAAC	960
15	AGTTGACACT CACTTATTGT AAAGTGAAGA TTGTTCTACT GCATGTGAAG TGGACCATGC	1020
	AGATTTCTGT ATGTTCTCAG TATGCATCAC TAGATAATAA AGTCTTTGT GAACAAGGCA	1080
	TTTGTAGCCA TTTTTAAAAAG TTTTGTCTT CAGTGCTGGT AAGTCAGGTA ACCATAAAAT	1140
20	AGTTAAAAGC AACCTTTGT TTTTTCTG AAAGTTTTA ATTGAAAGTA TTATTAGTTA	1200
	AAGATGTAAA CCTAGCCAAA ATTACCAAGTT TATTAATAAT TAGGATCCTA ATTATTICAA	1260
	AAAATCCTAC AAATATTGTC AGCTTCAGT GTAGTGAGAT TATTCTGTA GGTTATGGGG	1320
25	TATAATTCAAG GATTTAACTA ATGTTCTGC TATTTCTCA CTTTTCTTT TGATGGTGCG	1380
	GAAAGAGAAA AAGGAAAACG GGGCACAGGC CATTGACGC CTTCTCCAAG GGGTCTGATT	1440
30	TGCTGAGACA CCAGCTTCAC CTTCTTAACA AGGCACCTAA TTACAACAAG CATGCACATT	1500
	TTGGTGCATT CAAGAATGGA AAATCAGAAT AGCAGCATTG ATTCTCTGG TGCAGCTCAG	1560
	TGGAAGATGA TGACAACCAG AAGACATGAG CTAAGGGTAA GGGACTGTC TGAAGAACCT	1620
35	TTCCATTTAG TGATCAAGAT ATGGAAGCTG ATTCTGAAA ATGCTCAGTG TGTACTCTAA	1680
	TTATTTATGG TACCATTTGA ATTGTAACCT GCATTTAGC AGTGCATGTT TCTAATTGAC	1740
40	TTACTGGGAA ACTGAATAAA ATATGCCTCT TATTATCAA	1779

45 (2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2110 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

55	GCGGCCGCTG CAGCCCCGAG CTGAGCTAGC CGTCCGAGCC GAGCCGTCGG AGCCGGGGAA	60
	GCCGGCCGCTC GTGGCGGCCA GAGGAGAGGA GAGGCAGCAG CATGGCGAGT	120
60	GTCCCTGICCC GACGCCCTGG AAAGCGGTCC CTCCTGGAG CCCGGGTGTT GGGACCCAGT	180

	GCCTCGGAGG GGCTCGGCT GCCCCACCCCT CGGAGCCACT CCTAGAAGGG GCGGCTCCCC	240
	AGCCCTTCAC CACCTCTGAT GACACCCCT GCCAGGAGCA GCCAAGGAA GTCCCTTAAGG	300
5	CTCCCAGCAC CTGGGCCCTT CAGCAGGTGG CCTTTCAGCC TGGGCAGAAG GTTTATGTGT	360
	GGTACGGGGG TCAAGAGTGC ACAGGACTGG TGGGCCAGCA CAGCTGGATG GAGGGTCAGG	420
10	TGACCGTCTG GCTGCTGGAG CAGAAGCTGC AGGTCTGCTG CAGGGTGGAG GAGGTGTGGC	480
	TGGCAGAGCT GCAGGGCCCC TGTCcccAGG CACCACCCCT GGAGCCCGGA GCCCAGGCC	540
	TGGCCTACAG GCCCCTCTCC AGGAACATCG ATGTCCAAA GAGGAAGTCG GACGCATGGA	600
15	AATGGATGAG ATGATGGCGG CCATGGTGCT GACGTCCCTG TCCCTGCAGCC CTGTTGTACA	660
	GAGTCCTCCC GGGACCGAGG CCAACTCTC TGCTTCCCGT GCGGCTGCG ACCCATGGAA	720
20	GGAGAGTGGT GACATCTCGG ACAGCGGCAN CAGCACTACC AGCGTCACT GGAGTGGGAG	780
	CAGTGGTGTGTC TCCACCCCT CGCCCCCCCA CCCCCAGGCC AGCCCAAGT ATTTGGGGGA	840
	TGCTTTGGT TCTCCAAA CTGATCATGG CTTTGAGACC GATCTGACC CTTTCCTGCT	900
25	GGACGAACCA GCTCCACGAA AAAGAAAGAA CTCTGTGAAG GTGATGTACA AGTGCCTGTG	960
	GCCAAACTGT GGCAAAGTTC TGCGCTCCAT TGTGGGCATC AAACGACACG TCAAAGCCCT	1020
30	CCATCTGGGG GACACAGTGG ACTCTGATCA GTTCAAGCGG GAGGAGGATT TCTACTACAC	1080
	AGAGGTGCAG CTGAAGGAGG AATCTGCTGC TGCTGCTGCT GCTGCTGCCG CAGACCCCA	1140
	GTCCCTGGGA CTCCCCACCTC CGAGCCAGCT CCCACCCCA GCATGACTGG CCTGCCTCTG	1200
35	TCTGCTCTTC CACCACCTCT GCACAAAGCC CAGTCCTCCG GCCCAGAACAA TCCCTGGCCCG	1260
	GAGTOCTCCC TGCCCTCAGG GGCTCTCAGC AAGTCAGCTC CTGGGTCCCTT CTGGCACATT	1320
40	CAGGCAGATC ATGCATAACCA GGCTCTGCCA TCCCTCCAGA TCCCAGTCTC ACCACACATC	1380
	TACACCAGTG TCAGCTGGC TGCTGCCCTC TCCGCGCGCT GCTCTCTMC TCOGGTCCGG	1440
	AGCCGGTCGC TAAGCTCAG CGAAGCCCA GCAGCCAGCA CCTGCGATGA AATCTCATCT	1500
45	GATCGTCACT TCTCCACCCCG GGGCCAGAG TGGTGCCAGG AAAGCCCGAG GGGAGGCTAA	1560
	GAAGTGCAGC AAGTGTATGG CATCGAGCAC CGGGACCAGT GGTGCACGGC CTGCCGGTGG	1620
	AAGAAGGCCT GCGAGCGCTT TCTGGACTGA GCTGTCCTGC AGGTCTACT CTGTTCCCTGG	1680
50	CCCTGCCGGC AGCCACTGAC AAGAGGCCAG TGTGTACCA GCCCTCAGCA GAAACCGAAA	1740
	GAGAAAGAAC GGAAACACGG AGTTTGGCT CTGTTGGCTA AGGTGTAACA CTTAAAGCAA	1800
55	TTTTCTCCCA TTGTGCGAAC ATTATTTTTT TTAAAAAAA GAAACAAAAA TATTTTTCCC	1860
	CCTAAAATAG GAGAGAGCCA AAACTGACCA AGGCTATTCA GCAGTGAACC AGTGACCAAA	1920
60	GAATTAATTA CCCTCCGTTT CCCACATCCC CACTCTCTAG GGGATTAGCT TGTGCGTGTGTC	1980

AAAAGAAGGA ACAGCTCGTT CTGCTTCCTG CTGAGTCGGT GAATTCTTTC CTTCTAAAC	2040
TCTTCCAGAA AGGACTGTGA GCAAGATGAA TTTACTTTTC TTAAAAAAA AAAAAAAA	2100
5 AAAAACTCGA	2110

10 (2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 938 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

20 GGCACAGGAA AAAAAAGAAA AAAGAAAAAA GAAAAAAGTT TTGTACCCA CAGATTAGCA	60
TTTTCTTGAT GTTGAAGAA AGTTAACGCT ATGTCCTAAAT TTAAAAATGA GCACAAACTA	120
25 CTTAACAGAT GTCTGTTCCC TCTTCTCTTA CTTAAATTAT CTTTATTTTC ACCATCACCT	180
CCCAGTGCCG AACACCTGAN CTCTGTGTT TGTTGGTTGA TCCTGGGTTG CCAAGTCT	240
30 ATTGGTCAG TCCCTGGCCT GTGGGGCGGT CTCAGGAAGT GGCAATGCTCT TCAMGRAGGA	300
TCGTTCATYT CCAGTATAAC CAWTTGTTA ATAATAGTTG ATAATTCCCA GCTTTTACCA	360
GATGARTTTT GACTTATTTT TCCTCCCTTG ACCTGTTCAA AGCTAACATA TCTCGGTCA	420
35 TTCGGAGAGG GTGGGGGATT TGAGAATGTG AGGAGGAGTG GGGTTAGAAT GGGTTTGCCT	480
ATCTGGCAA GGAAAGAGTT CCTAGTCGAT TGGGCACAAT GACAAAAATGA TTCCATGGAT	540
AGAATCGTCC CATGTTGCTG GAACACCTCA CGTGTGTTGA ACCCTTAAA TTCCCTGCCAT	600
40 CCCCTCTCTG ATTCCCCACC TCCCTGTAGT TTCCACAGGA TTTATCTCTC TGTACCCCCG	660
TCCTOCAAAT CTACTCTGTC AGCCTCTCT CCATCCCTTA CTTCCCTCT AAATTCCAGG	720
45 AGATGACCTC ACTTTGCAA GCAAATTGGA GCCACCAAAT TGTAGCTCTC CTCGGTGGAA	780
ACTGCATCTG TGCTCATCCC TGCACCTCT TGCAAGAAAGC CGCCCCCTCA GGCAAGATG	840
AGTGCCTGGC CCCCATGGGA GACTCAGACA CTTTGACCCC TTGTGACTTC AGCATCTCCC	900
50 TCTTTAAAGA TTCTCTCCCA ACATTCAGTC GTGCTCGA	938

55 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1551 base pairs
 (B) TYPE: nucleic acid
 60 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

5	AGGCTGGACT AAGCATAGAG AACCAGGAGA GAAAGAAAAGA TTTAAGAGAC TGAGTAATAT	50
	TTTTGACAG ATCAATTAAAG AAACGTGAGTA ATTTTTTTTT TCTCCAAAAG GGCATGGGTT	120
10	TTTTTTTGT TTTGTTTTTG CTCATTTGG CACTTTCTAG GGATTGGTCT ATAAATTTT	130
	TGAAAGATCA TAGGATAAT TTCTTTGTAG CAACTTCCTA TTTTAGTGT TATGTTAGGG	240
	GARCCCCARG TGTCCCTGCT GATAGCCAT TAGGCCACT TCTCAGCCTC TGGCTACATC	300
15	ATAATGCTTT TTTTCTATC TTGCAAAGT TTCCMAAAAA TTKAKGTTTT CTAATTTTAA	360
	AAAAATTGGT TGTGGAGATG GGATGGGACC TCTTATAAG CCCTGAAAAT AAGTGTGTTN	420
20	TTTTAAGTGC TATTCTGCTA TAAACCTGAT TCTCACTTTT TTCTGTAGAC AACAGTTTT	430
	TATAATATAT CTATTTGTG TGGACATTAT TTCCCTTTAA CCAAACTGA AATTCCATAG	540
	TGTAWACMTT CTCCACATTT TCTTGTGATTA ATACTTYCTT AAAATAGACA CTGGATTGG	600
25	CACCAGCTGT CACCAATAAA GCTGCCCTGA ACATTGTCAA TCAATCCTGT TARCCAATT	660
	GAGAATTTTT CTGGAATGCT TAGTTAGGGA TGAAATTGCT GGGTTATAGG TATGAGTATG	720
30	CTTGATATAC TTTTCTCCAG AATGTCTACA CCTGTGTGTA CACCACATCT CCAGAGATAG	730
	GGGAATCTTA TGTCCCTGCT AACIGCTCTC GTTATTAAAT TTCTGACAT TTGCCGCCGC	840
	CGCCGCCCGG TGCCCCAAC ACACACATGG TATAAAGTGG TAGTTCTTG TTTAAATTG	930
35	AACTTTGAA TGATTTGAAT TTGGCATT TTTGTATCC TGAGTTATT TGGTTCCCG	950
	TTATGTGAAT ATCCCTTCC TATGTTTAA CTACTTTCT AATTGTCCC TTTTTNGGT	1020
40	TATCAAATTG CAGGCCATTG TCTATTCCAT CGTCACTTTT GGGTATTGGA AACATCTTC	1030
	CATTCTGTAG CCTGTCTGTT GAACATAAT CTTGATTTT ATGTAATCAG ATTTCCTCC	1140
	TTACGGTTAT GTTCTGGAA TTTTATTAA GAAATCTTT TCTATCCTGA GACCACAAA	1200
45	ATGTCCCCAC CATTTCCTC TGTTTCATAG TTTTGCCTTG TATTTAAAT CCTTTAAGGC	1260
	ATGTGTAGTT CATTATAT GGTGTGAAAT AGTTCTTATT CATTATTCA ACACATATTG	1320
50	GTGGAGTGCC TGCTGATGGT AGTACTCTTC AGAGTACTTT GTATATATTT GTGAACACAT	1330
	ATTCTTGCCC TGGAAGCTTA TGTTGTCNTT CAAGGTAGAT CCNTACTCGG TTTCCACCTG	1440
	TTTTCTTCAG CCCTCAGGAT GAATTCCACA ATTTTACACA TAGCACCAGT TAAGGAATAG	1500
55	GCTTTATTGG AGAAAAGGAA GGCTTATTAG ACCAGCATCA GCAAAAAAAA A	1551

60 (2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 997 base pairs
- (B) TYPE: nucleic acid
- 5 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

10	AGAGAGTCCT CAACAGAACCC TAATCATGCT GGCACCCCAA CCTCGTACTT CTAGCCTCCA	50
	GAAGTGAGAG AACATAAAACT CCAGTTGTCTT AAGGCTACCCCA CCTCGTGGTA TTTCGTATTA	120
	TAGCCCAAGC TAAGTCAGGT GGAAAGGCCAG AAATATTTCG AGAGAGTCGA TTTCCTACAA	180
15	AACAGAGTTG TTCTAAATGA AATGGCCAGA TATTCATCT CCTCGTACTT AGTATTTCG	240
	AAAGTTTCAT TAAACACCAC TTGGCCAGCA CCCAGGCCTG CCTCGTCTCG AACGGCAAC	300
20	AAAAGCAAAT GATTTGAGGA ACAAAAGAGT GGACACAGAG CCTCTCAGAA GATGGCTCCA	360
	TCTTCTGAGA TGATCTCTG AGATCATCCTA TTTTCTGCCG CCTCGTCTCT AGTCCAACTG	420
	TAGTAGATAA GAGCAAAGAC ACTTCCTGAT CCTGTGGAAA ATGCTGCCG CCTGCTGATG	480
25	GAGAGGCTGA CACTGGGACC AACAGAAGGC CGGACATTA TTTCGTGGAG CCTCTCTGAA	540
	CCTGGGCCT CCTCAGGCCT TGTACCTTGC ACTCCCCATG CCTCTGAGC ACCTGGTACG	600
30	CTGAAGTTAG GTATTTGAAG AGATAATTG CCCCCAACCA AGATTTCTT AAAGCAAA	660
	GGAAACCACT AAATTCCACT TGACAAACCA GTTTGTTGAG TTTTACTTT TGCAAAATTG	720
	AAACTTTCTC TTTGGCACCA TATGATTCTG TTACATTAGG CCTCGTCAAT CCTAAGATAAC	780
35	ACAGCTAGGT CTACCACTG CCAGTGGTCA AGATGAAAG CCTCGTCTCG AGAGAGATCA	840
	GTTTCTAATA ACCTAACAGT TTTCCTTGGS TATTACCAA AAAAAGAAA TTAGAATCAA	900
40	ATGTCAGTGC CATGCCAGCA AGTACAGATA TGGAATGAA AGCTTGCTT ACAACTGCCA	960
	GATTTGTTTG TTAATAAAAT TGATTTGGAT CACTCGA	997

45

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1496 base pairs
- (B) TYPE: nucleic acid
- 50 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

	GAATTCGGCA CGAGTGACCA CAGATATCTT TGGCTTTCAG CCTCGCCCAA ATGCTGMCMA	60
	CTATGTTTTT TTAAATCGAT TGACATCTCA TGAAATCCACA ATTTCGCGG CCTTTCCATC	120

60

	TTTCCATCT TTGTCATAGC TTCAATCACGC ACGATGGAGG TCACTTCAGC ACTATCCGGA	180
	GCGGCCCTCAC CGACAGATCR GTGAATTTC CTTTCCCTTT TCCTGATGTA CCGGATTGTC	240
5	GACTCGTTAA CATTGAGCTC ATGGCCAACA GCACTGTAAC TCATGCCTGA TTGGAGCTTA	300
	TCCAACACGC GGAMTTCTC CGTAAGGSAM ATCAMGGTCT TCTTCGCTT AGGAACACTG	360
10	GGCARARCTT AARCACTACG CTTGGGGCC ATTTTAGAAA GCAAAACCAC CCACAAAAAG	420
	CAGAAAAAAA AGTGTCAAGTA AACAGACTGN NGANAGGACT CTTTGTTTAC AGCACAGGAG	480
	CTGGCACTAG AAGGCGGCAG TTCTCCCCAG TTCAAACCTTC AGCTGGAAC CTTACCTCCG	540
15	CCAACCTCAA ATTTCACCC TCTGCGCATG CCCGGGAAAS AAACCCCCAG AACAGTACCG	600
	TGATGATTGA TTTTAGGGTT ACAAATACAT TTTAGCAAGT AAGTGAATTG GGCATTACGA	660
20	ATTAATGATT AATGAAGGTC ACCTGTATTT CCATAGATAT GTAATTAT TTAAGCAGGT	720
	TTATTATATT AAGGCGGSGA GGCAGCGCCG AAGACTACAA GTTCCAGCAT GCACCGCGTC	780
	CGGGCGGGTT CGGGCTCCCA GCGAGGGCTT CAGGGACGCC AGCCCCGGAGG CATGGGCCGG	840
25	AAGTGTCAAGTA GGGCAACCCAC GTAGTACTCTC CTGGCGATGT GCAAAGCGCT GTGGGGGGCC	900
	GCCCTAGCTG CGCTCGCCGC CGCCGGGGCT CTATGGTCTC TCCCTAGAGC TTTGCCGTIG	960
	GAGGCGGCTG CTGGGGTCTT GTGAGTTTGA CCAGCGTCGA GCCCAGCAA CATGGAGGAA	1020
30	TTCGACTCCG AAGACTCTC TACGTGGAG GAGGACGAGG ACTACGTGCC GTGGGGTGAG	1080
	CGATTCCGCC TGAGGCGAGA AGCGAATTGC CCCGCCCCAC GCCTCACGTG AGGCGCGCTC	1140
35	TGCCCGCGCG GGCCTCTGCC CTGTGGCCCA GGTGGTCCAG GGGGGCTCCT GTTCTCGAGC	1200
	GTCCGCTCCC TCAGGCCCCCT CATQCTCGGC CGCTCCGCC CGAGGCGTGT GCGCGTGGCG	1260
40	GTTCTGTGCT CCCCTCCCCGT TGGGCAGCTC CGGCCGCCGC CCCCTCTTGC AGCGCGGGAA	1320
	CGGCACATGG ACACGGCCCG TTGTCGCTAG GGACGCTCGT CGGTCAAGCCC CGAACGACAA	1380
	CGCTGCTTCA GAAGTCGGGG CGGCAGTTCG AGCCTTGAA GTTTTTTCA GCCCTGGCCC	1440
45	GAGAGAGCTG CTGGCCAACA ACCCGTCCAA GATAGAGCTG TCCGNTCTCC GNCTGG	1496

50 (2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1308 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

60 TTGGCANCNG GGAGAGGGAA AGAGGAGGAA ATGGGGTTTG AGGACCATGG CTTACCTTC

	CTGCCTTGAGCCATCACACCCCATTCTT CCTCTTTCCCC TCTCCCCGCT GCCAAAAAAA	120
5	AAAAAAAAAGG AAACGTITAT CATGAATCAA CAGGGTTTCA GTCCCTATCA AAGAGAGATG	180
	TGGAAAGAGC TAAAGAAACC ACCCTTTGTT CCCAACTCCA CTTTACCCAT ATTTTATGCA	240
	ACACAAACAC TGTCCTTTG GGTCCCTTTC TTACAGATGG ACCTCTTGAG AAGAATTATC	300
10	GTATTCCACG TTTTAGGCC TCAGGTACC AAGATAAATA TATGTATATA TAACCTTTAT	360
	TATTGCTATA TCTTGTGGA TAATACATTG AGGTGGTGCT CGGTGATTTA TTATAATCTG	420
	AACCTAGGTA TATCCTTTGG TCTTCCACAG TCATGTTGAG GTGGGCTCCC TGGPATGGTA	480
15	AAAAGCCAGG TATAATGTAA CTTCACCCCCA GCCTTTGTAC TAAGCTCTTG ATAGTGGATA	540
	TACTCTTTA AGTTTAGCCC CAATATAGGG TAATGGAAT TTCCCTGCCCT CTGGGTTCCC	600
20	CATTTTTACT ATTAAGAAGA CCAGTGATAA TTTAATAATG CCACCAACTC TGGCTTAGTT	660
	AAGTGAGAGT GTGAACTGTG TGGCAAGAGA GCCTCACACC TCACTAGGTG CAGAGAGCCC	720
	AGGCCCTTATG TTAAAATCAT GCACTTGAAA AGCAAACCTT AATCTGCAA GACAGCAGCA	780
25	AGCATTATAC GGTCATCTTG AATGATCCCT TTGAAATTTT TTTTTGTTT GTTTGTTAA	840
	ATCAAGCCTG AGGCTGGTGA ACAGTAGCTA CACACCCATA TTGTGTTGTC TGTGAATGCT	900
30	AGCTCTCTTG AATTTGGATA TTGGTTATTT TTTATAGAGT GTAAACCAAG TTTTATATTC	960
	TGCAATGCGA ACAGGTACCT ATCTGTTCT AAATAAAACT GTTTACATTG ATTATGGGT	1020
	ATGTATGACC TTCATTTTCC AAGAAATAGA ACTCTAGCTT AGAATTATGG ATGCTCTAAA	1080
35	ATGTCAGAAT GGGAACTCTC CTGGAAGTTC TCCCAAACTC AGAGACAGCA CTGCCTTCTC	1140
	CTAAATGATT ATTCTTTCTC CCCTGTTTTC TGGTATTTTC TAGGCATCCT TCTCACCACA	1200
40	GCCATAACCC TTTTTTACTT CCATTAGGCC GTATAACTGG NGGGACNGCT GGTCGGTATA	1260
	TAATACTGGT WCCAACAMAG GGGTCTGGA TGTACACMAG GTTATCTT	1308

45

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1705 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDENESS: double
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

	TGGCCATGGA AGCGCTAGAA GGTTAGATT TTGAAACAGC AAAGAAGGAT TTCCCTGGAT	60
	CTGGAGACCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA	120

60

	TCAAGGAGCC CAAAGCCGCC GTGGAGATGT ACATCTCAGC AGGAGACAC GTCAAGGCCA	180
	TCGAGATCTG TGGTGACCAC GGCTGGTGTG ACATGTTGAT CGACATGCC CGCAAACCTGG	240
5	ACAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGCTACCTA CCTCAAGAAC CTGGACAGCC	300
	CTGGCTATGC TGCTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCCCTG GTGCAGCTGC	360
10	AGTGGAGACC CAGCGCTGGG ATGAGGCCTT TGCTTTGGGT GAGAACATC CTGAGTTAA	420
	GGATGACATC TACATGCCGT ATGCTCAGTG GCTAGCAGAG AACGATCGCT TTGAGGAAGC	480
	CCAGAAAGCG TTCCACAAAGG CTGGCGACA GAGAGAACCG GTCCAGGTGC TGGAGCAGCT	540
15	CACAAACAAAT GCGGTGGCGG AGACGAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT	600
	GTCCATGCAG TGCCTCGATA TAGCTCAAGA TCCTGCCAG AAGGACACAA TGCTTGGCAA	660
20	GTTCTACCAC TTCCAGCGTT TGGCAGAGCT GTACCATGGT TACCATGCCA TCCATGCCA	720
	CACGGAAAGAT CCGITCAGTG TCCATCGTCC TGAAACTCTT TTCAACATCT CCAGGTTCC	780
	GCTGCACAGC CTGCCAAGG ACACCCCCCTC GGGCATCTCT AAAGTGAAAA TACTCTTCAC	840
25	CTTGGCCAAG CAGAGCAAGG CCCTCGGTGC CTACAGGCTG GCCCGGCACG CCTATGACAA	900
	GCTGGCGTGGC CTGTACATCC CTGCCAGATT CCAAAAGTCC ATTGAGCTGG GTACCCCTGAC	960
30	CATCCCGGCC AAGCCCTTCC ACCACAGTGA GGAGTTGGTG CCCCTGTGCT ACCGCTGCTC	1020
	CACCAACAAAC CCGCTGCTCA ACAACCTGGG CAACTGCTGC ATCAACTGCC GCCAGCCCTT	1080
	CATCTCTCC GCCTCTTCCT ACGACGTGCT ACACCTGGTT GAGTTCTACC TGGAGGAAGG	1140
35	GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGGAG GTGCTGAGAC CCAAGCGGGA	1200
	TGACAGACAG CTAGAGATTT GCAAACAAACA GCTCCCAGAT TCTTGCGGCT AGTGGGAGAC	1260
	CAAGGGACTC CATCGGAGAT NAGGACCCGT TCACAGCTAA GCTRAGCTTT GAGCAAGGTG	1320
40	GCTCARAGTT CGTGCAGTG GTGGTGAGCC GGCTGGTGCT GCGCTCCATG AGCCGCCGGG	1380
	ATGTCTTCAT CAACCGATGG CCCCCACCCC TGAGGTGGCA ATACTTCCGC TCACTGCTGC	1440
45	CTGACGCCCTC CATTACCATG TGCCCTCCT GCTTCCAGAT GTTCCATTCT GAGGACTATG	1500
	AGTTGCTGGT GCTTCAGCAT GGCTGCTGCC CCTACTGCCG CAGGTGCAAG GATGACCCCTG	1560
	GCCCATGACC AGCATCCTGG GGACGGCCTG CACCCCTCTGC CGCCCTTGGG GTCTGCTGGG	1620
50	CTGTGAAGGA GAATAAAAGAG TTAAACTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1680
	AAAAAAAAAA AAAAAAAAAA AAANA	1705

- (A) LENGTH: 999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

	AGCAAATCAC CTTAACGATC TGGAAATGAAA CTGTGACCAAG TGCCGCCCTG GGTGGTTC	60
10	GAGAGACTGC CGTCTTCTTG TTTGCCATA GGTGCTGGGG CCCCGGCTTC AGTCACTGTC	120
	TCAGACAGKA GTCCCGATAA GCAGATCACC AGTCCTCCAC TGTCCTTCCT GTCGGCC	180
15	CTGCATGAGA AGATAGCTGC TTCCCTCCCTC TTTTCCCTACA CTGTAAATTAA TTGTTTTACA	240
	ATTGAGTGYC TTAATAATAG TYTACAAATA CTATGTATTT ATGCAAAACT GTTAAAGTTC	300
	TCATCTGTTA TGATTGGATA CTIIGTCTTG TCAGTAGTGG TCAGCATTGG GTTGTGAGCT	360
20	TGTCCTACTC CATACTGTT TATCCTGCTA TGCAATTAC ATTGTGTGTT CACATCTATT	420
	CCAAGGAGCC TTGCTAGAAA CAACACTGGC GGTCCTGCA GGCCAGGCAG GCATTGGCC	480
	ATGCTGTGTC CCATAGGAGC CAATGGAAAG AACGTAGCTT GGTCTGCTAG CCAGCCGTGG	540
25	GGTGGCGCAG GCCAGGCAGC CTCTGCACCA GAGTCCAGCA CCTGCCATT CCCAGTCAC	600
	ACAATCATAAC TCTTCTTCA TAGAGATTTT ATTACCACCT AGACCACCT AGTTTCTTC	660
30	TCTGTTAGTG TCCGTAGCTC TTTTCAACA AAATGTAGGT ACAGACCAAT CCCTGTCCC	720
	TCCCCAATCA GGAGCTCCAC ACCATGAGTT GTTGGTTTT CCAGAAGCTG CCAGTGGGTT	780
	CCCGTGAATT GCGTTAAGAT ATCGATGATK TTTTTTATTG TTTTCTTCT TGTAAAAA	840
35	AATAATATAT TTAAAGGCAG TATCTTTGT ACTGTGAATT TGCAGTAGAA GATGCAGAAT	900
	GCACTTTTTT TTTACTCTG TTGGTGTGTA TTGTATATAG TGTGTGTGCT TCTTGTGATG	960
40	AAAAATAAAACT TTTTCTTAT AAAAAAAA AAAAAAAAC	999

45 (2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

55	GGCACGAGTA GCATTCATT TAATCTGCAG GTATATTCTC CCAACAGTT ATTGTCA	60
	GATGTCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA	120
60	GCCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA	180

	TGTCCCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCGGCTAGCG GTTTGAGCCA	240
	GAGAAATGACA GCTCTGGTTT GGAGAAAAGG GCCGGATGGT GGCTCTAGAA AGCCCATCCT	300
5	TCTGCTCTTC TTTTTCTCC CCCTTATATT GTGCTTCAT TCATTCAATC ATTCAATCAA	360
	CATTGTTGA GCACCTATTA TGTGTCAAGC TCTGTGCTAG CCTCTGGAAA ACCTGCCCTC	420
10	ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGGTTGTCA	480
	GGGTCTCACA GACCAGTGGC CCCTCATCCA GACCGATGAG GTCAAAGAAC GCATCCAGGC	540
	GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC	600
15	TGAAGGTGGC AGTGCCTGGA GTCTTGATTC CAGCAGAGGG AGAGCAGTCT GTGAAAAGGC	660
	ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACCTCA GGTAGTTCTG GATGGCCTG	720
20	GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA	780
	TCCCAATAAA CCCATTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTAA GGCTGGTCCG	840
	AGTAGAATGA TTTTACAAC GAATTGATCA CAACCAGTTA CAGATGTCTT TGTTCCCTCT	900
25	CCACTCCCAC TGCTTCACCT GACTAGCCTT TAAAAAAAAA A	941

30 (2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

40	TAAGTGAAT CCCCCGGGGT TGCAGGGAAT TCGGCACGAG GCATTCTGAG AAGCTTAAGA	60
	CATACTTTGA AGACAACCT AGGGACCTCC AGCTGCTGCG GCATGACCTA CCTTTGCACC	120
	CCGCAGTGGT GAAGCCCCAC CTGGGCCATG TTCTTGACTA CCTGGTTCTT CCTGCTCTCC	180
45	GTGGCCCTGGT RCGCCCTCAC AAGAACCGGA AGAACGCTGTC TTCTCTTGT AGGAAGGCCA	240
	AGAGAGCAA GTCCCAGAAC CCACTGCGCA GCTTCAAGCA CAAAGGAAAG AAATTCAAGAC	300
50	CCACAGCCAA GCCCTCTGA GGTGGTTGGG CCTCTCTGGA GCTGAGGACA TTCTGGAGCA	360
	CAGGCTTACA CCCTTCCGTGG ACAGGCGAGG CTCTGGTGCT TACTGCACAG CCTGAACAGA	420
	CAGTCTGGG GCCGGCAGTG CTGGGCCCTT TAGCTCTTGC GCACTTCCAA CCTGGCATCT	480
55	TGCCCTTGA CAACAGAATA AAAATTTAG CTGCCCAAA AAAAAAAAAA AAAAAAAAAA	540
	CTCGAGGGGG GGCCCGTACC CAATTGCCCC TATAA	575

(2) INFORMATION FOR SEQ ID NO: 220:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3018 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

GCCAGCCTTA	CAGGTTTAC	GTGAAATGAA	AGCCATTGGA	ATAGAACCT	CGCTTGCAAC	60	
15	ATATCACCAT	ATTATTCCCC	TGTTTGATCA	ACCTGGAGAC	CCTTTAAAGA	GATCATCCCT	120
	CATCATTTAT	GATATAATGA	ATGAATTAAT	GGGAAAGAGA	TTTTCTCAA	AGGACCOGGA	180
20	TGATGATAAG	TTTTTCAGT	CAGCCATGAG	CATATGCTCA	TCTCTCAGAG	ATCTAGAACT	240
	TGCCTACCAA	GTACATGCC	TTTTAAAAAC	CGGAGACAAAC	TGGAAATTCA	TTGGACCTGA	300
25	TCAACATCGT	AATTTCTATT	ATTCCAAGTT	CTTCGATTTG	ATTTGTCTAA	TGGAACAAAT	360
	TGATGTTACC	TTGAAGTGGT	ATGAGGACCT	GATACCTTCA	GCCTACTTTC	CCCACTOCCA	420
30	AACAATGATA	CATCTCTCC	AAGCATTGGA	TGTGGCCAAT	CGGCTAGAAG	TGATTCTAA	480
	AATTGGGAA	AGATAGTAAA	GAATATGGTC	ATACTTTCG	CAGTGACCTG	AGAGAAGAGA	540
35	TCCTGATGCT	CATGGCAAGG	GACAAGCACC	CACCAGAGCT	TCAGGTGGCA	TTTGCTGACT	600
	GTGCTGCTGA	TATCAAATCT	GCGTATGAAA	GCCAACCCAT	CAGACAGACT	GTCAGGATT	660
40	GGCCAGCCAC	CTCTCTCAAC	TGTATAGCTA	TCCTCTTTT	AAGGGCTGGG	AGAACTCAGG	720
	AAGCCTGGAA	AATGTTGGGG	CTTTTCAGGA	ACCATAATAA	GATTCTAGA	AGTGAGTTGC	780
45	TGAATGAGCT	TATGGACAGT	GCAAAAGTGT	CTAACAGCCC	TTCCCAGGCC	ATTGAAGTAG	840
	TAGAGCTGGC	AAGTGCCTTC	AGCTTACCTA	TTTGTGAGGG	CCTCACCCAG	AGAGTAATGA	900
50	GTGATTTGCA	AATCAACCAG	GAACAAAAGG	AAGCCTAAG	TAATCTAATC	GCATTGACCA	960
	GTGACAGTGA	TACTGACAGC	AGCAGTGACA	GCGACAGTGA	CACCACTGAA	GGCAAATGAA	1020
55	AGTGGAGATT	CAGGAGCAGC	AATGGTCTCA	CCATAGCTGC	TGGAATCACA	CCTGAGAACT	1080
	GAGATATACC	AATAATTAAAC	ATTGTTACAA	AGAAGAAAAG	ATACAGATTT	GGTGAATTG	1140
60	TTACTGTGAG	GTACAGTCAG	TACACAGCTG	ACTTATGTAG	ATTTAAGCTG	CTAATATGCT	1200
	ACTTAACCAT	CTATTAATGC	ACCATTAAG	GCTTAGCATT	TAAGTAGCAA	CATTGCGGTT	1260
65	TTCAGACACA	TGGTGAGGTC	CATGGCTCTT	GTCATCAGGA	TAAGCCTGCA	CACCTAGAGT	1320
	GTCGGTGAGC	TGACCTCACG	ATGCTGCTCT	CGTGGGATTG	CCCTCTCCTG	CTGCTGGACT	1380
70	TCTGCCCTTG	TTGGCCTGAT	GTGCTGCTGT	GATGCTGGTC	CTTCATCTTA	GGTGTTCATG	1440

	CAGTTCTAAC ACAGTTGGGG TTGGGTCAAT AGTTTCCCAA TTTCAAGGATA TTTCGATGTC	1500
	AGAAATAACG CATCTTAGGA ATGACTAAAC AAGATAATGG CAGTTAGGC TGCACAACAG	1560
5	GTAAAATGAC TGTAGATAAA TGTTGTAATT AGTGTACACG TTTGTATTTTG TGTTAATATA	1620
	GCCGCTGCCA TAGTTTCTA ACTTGAACAG CCATGAATGT TTCATGTCTC CCTTTTTTT	1680
10	TTGTCTATAG CTGTTACCTA TTTTAGTGGT TGAAATGAGA GCTAGTGATG ACAGAAGGAT	1740
	GTGGAATGTC TTCTTGACAT CATTGIGTAT TGCTGGTAAT CAAGTTGGTA ACGACTACTT	1800
	CTAGCAGCTC TTACCACTAT GACTTAAGTG GTCCTGGAAG GCAGTAAGTG GAGGTTTGCA	1860
15	GCATTCCTGC CTTCATGAGG GCTTCTACCA CTGACCACCT TGACAGTACCCAG	1920
	ATTTACTTAG GTACCCCCACG AGTCGTCCAC ATAAGCAGCT TCATCTTAC CTTGCCAGAG	1980
20	TTGACAATTAA TGGGATACTC TAGTCTACTT ATACTTGTGT TCCCACATCTGT CTGCCATCCT	2040
	CTGAAGGCCA GGACCCAGTC ATACATCCTT AGAAACAAAA GTATGGTTT TGTTTTCTCT	2100
	TGGAATGTCA GGTCTTAAGG CATTAAATTG AGGGACAAAA AAAAAAAA GCCGATATAG	2160
25	TAGCTAGCTA CTTAACGATC CATGGTATT GCTCCATATC AAAGCAGATT TCCAGGACAG	2220
	AAAGAGTAAA TTAGCCCTCA GTCCTGGTTT ACAGCTTCCA AGGAGAGCCT TGGSCACCTG	2280
30	AAATGTTAAC TCGGTCCCTT CCTGTCCTCA GTTCATCAGC ACCTGCAGAT GCCTGACTCT	2340
	TGTTAGCCTT ACTATTCAAT ACAGTCCTTA GATTCACCGT ATGCCCTCTTC CTATCCAGGC	2400
	ACCTATTCTG AATCACCATG TTGCTCTGCA CCTAGAGTG ATAGGAGAAA ATCCATTGG	2460
35	GTAGATGGCC TATGAATTG TAGTAGACTT TCAAAATGAG TGATTGTTA GCTTGGTACT	2520
	TTAAAGTTG TGGTACAGAT CCTCCAAACC CATACTCTGA GCAATTAAC GCCTGAAACA	2580
	TAGAGAAAAA TTAAGGCCTC ACAGGATGAG TCTCCATTCT CTGIAAATGC TTATTTTATC	2640
40	ATAGTCTTTA GCCTCTAACT ATGAGTAAAA TGTTCTCTTC GGCCGGGTGT GGTGACTCAC	2700
	ACCTGTAACC TCAGCACTTT GGGAGGCAGA GGTGGGAGGA TCACCTAGGT CCAGGAGTC	2760
45	GAGACTAGCC TGGGCAACAT AGTGAGACAC CGGATCTACA AAAAAATAAA AAGCCAGACT	2820
	GGTGGTATGT ATCTGTGTCC CAGCTAATTG GGAGGGTGAG ATGGGAGGAT TGTTTGAGCC	2880
50	TAGGAGAGGG AGGTTGCAGT GAGCCGTGAT CGCACCACTG CACTCCAGCC TGGGCAACAG	2940
	AGCAAGACCC TGTCTTGGAG AAACCAGAAT TTGGAAGAG CAAATGGGGC TGAGTGCAGT	3000
	GGCTCATGCC TGTAATCC	3018
55		

(2) INFORMATION FOR SEQ ID NO: 221:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 968 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

GGCACGAGGG CCGCGGGACA TCCACGGGGC GCGACTGACA CGCGGGAGGG AGAGCAGTGT	60
10 TCTGCTGGAG CCGATGCCAA AAACCATGCA TTTCCTTATTC AGATTCAATTG TTTTCCTTTA	120
TCTGTGGGC CTTTTTACTG CTCAGAGACA AAAGAAAGAG GAGAGCACCG AAGAAGTGAA	180
15 AATAGAAGTT TTGCATCGTC CAGAAAATG CTCTAAGACA AGCAAGAAGG GAGACCTACT	240
NAAATGCCCA TTATGACGGC TACCTGGCTA AAGACGGCTC GAAATTCTAC TGCAGCCGGA	300
CACAAAATGA AGGCCACCCC AAATGGTTTG TTCTTGGTGT TGGGCAAGTC ATAAAAGGCC	360
20 TAGACATTGC TATGACAGAT ATGTGCCCTG GAGAAAAGCG AAAAGTAGTT ATACCCCCTT	420
CATTIGCATA CGGAAAGGAA GGCTATGCAG AAGGCAAGAT TCCACCCGAT CCTACATTGA	480
25 TTTTGAGAT TGAACCTTAT GCTGTGACCA AAGGACCACG GAGCATTGAG ACATTTAAC	540
AAATAGACAT GGACAATGAC AGGCAGCTCT CTAAAGCCGA GATAAACCTC TACTTGCAAA	600
GGGAATTGGA AAAAGATGAG AAGCCACGTG ACAAGTCATA TCAGGATGCA GTTTTGAAG	660
30 ATATTTTTAA GAAGAATGAC CATGATGGTG ATGGCTTCAT TTCTCCCAAG GAATACAATG	720
TATACCAACA CGATGAACTA TAGCATATTT GTATTTCTAC TTTTTTTTT TAGCTATTAA	780
35 CTGTACTTTA TGTAIWAAAC AAAGTCMCTT TTCTCCMAGT TGKATTTGCT ATTTTCCCCC	840
TATGAGAAGA TATTTTGATC TCCCCAATAC ATTGATTTTG GTATAATAAA TGTGAGGCTG	900
TTTTGCAAAC TTAAAAAAA ATTTAAAAAA ACTGGAGGGG GGCCCGTACC CAANTCGCCG	960
40 NATATGAT	968

45 (2) INFORMATION FOR SEQ ID NO: 222:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1404 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

55 CGTTTTCCGG CGGTGCGTTT GTGGCCGTCC GGCCCTCCCTG ACATGCAGCC CTCTGGACCC	60
CGAGGTTGGA CCCTACTGTG ACACACCTAC CATGCGGACA CTCTCAACC TCCTCTGGCT	120
60 TGCAGCCCTG TTCACACTAC CCTGTCAAAG TCAGATGCCA AAAAAGCCGC	180

	CTCAAAGACG CTGCTGGAGA AGAGTCAGTT TTCAGATAAG CCGGTGCAAG ACCGGGGTTT	240
	GGTGGTGACG GACCTCAAAG CTGAGAGTGT GGTTCTTGAG CATGCAGCT ACTGCTCGC	300
5	AAAGGCCCGG GACAGACACT TTGCTGGGA TGTACTGGC TATGTCACTC CATGGAACAG	360
	CCATGGCTAC GATGTCACCA AGGTCTTGG GAGCAAGTTC ACACAGATCT CACCGTCTG	420
10	GCTGCAGCTG AAGAGACGTG GCCGTGAGAT GTTTGAGGTC ACGGGCCTCC ACGACGTGGA	480
	CCAAGGGTGG ATGCGAGCTG TCAGGAAGCA TGCCAAGGC CTGCACATAG TGCTCGGCT	540
	CCTGTTGAG GACTGGACTT ACGATGATT CCAGAACGTC TTAGACAGTG AGGATGAGAT	600
15	AGAGGAGCTG AGCAAGACCG TGGTCCAGGT GCAAAGAAC CAGCATTCG ATGGCTTCGT	660
	GGTGGAGGTC TGGAAACCAGC TGCTAAGCCA GAAGCGCGTG GGCTCATCC ACATGCTCAC	720
20	CCACTTGGCC GAGGCTCTGC ACCAGGCCCG GCTGCTGGCC CTCCCTGGTCA TCCCGCCTGC	780
	CATCACCCCC GGGACCGACC AGCTGGGCAT GTTCACGCAC AAGGAGTTG AGCAGCTGGC	840
	CCCCGTGCTG GATGGTTTCA GCCTCATGAC CTACGACTAC TCTACAGGC ATCAGCCTGG	900
25	CCCTTAATGCA CCCCTGTCTT GGGTTGAGC CTGCGTCCAG GTCCCTGGACC CGAAGTCCAA	960
	GTGGCGAAGC AAAATCCTCC TGGGGCTCAA CTTCTATGGT ATGGACTACG CGACCTCCAA	1020
30	GGATGCCCGT GACCTGTTG TCGGGGCCAG GTACATCCAG AACTGAAGG ACCACAGGCC	1080
	CCGGATGGTG TGGGACAGCC AGGYCTCAGA GCACCTCTTC GAGTACAAGA AGAGCCGAG	1140
	TGGGAGGCAC GTCGTCITCT ACCCAACCTT GAAGTCCCTG CAGGTGCGGC TGGAGCTGGC	1200
35	COGGGAGCTG GGCGTTGGGG TCTCTATCTG GGAGCTGGCC AGGGCCTGGA CTACTCTAC	1260
	GACCTGCTCT AGGTGGGCAT TGCCTGCCTCC GCGGTGGACG TGTCTTTTC TAAGCCATGG	1320
	AGTGAGTGAG CAGGTGTGAA ATACAGGCCT NCACTCCGTT TGCTGTGAAA AAAAAAAA	1380
40	AAAAAAAAA AAAAAAAA AAAA	1404

45

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 707 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

55	NGCGCGCCTG CAGTCGACAC TAGTGGATCC AAAGAATTCTG GCACGAGGGC AGGTCCAGGG	60
	CTCAGAAATC AGCTCTATCTG ACGAATTCTG CGCAAGTTC CGCCTGGACT GCCCGCTGGC	120
60	CATGGAGCGG ATCAAGGAGG ACCGGCCCAT CACCATCAAG GACGACAAGG GCAACCTCAA	180

CCGCTGCATC GCAGACGTGG TCTCGCTTTC GATCACGGTC ATGGACAGC TGGCCCTGGA 240
 5 GATCCGGGCC ATGGATGAGA TCCAGCCCGA CCTGGAGAG CTGATGAGA CGATGACOG 300
 CATGAGCCAC CTCCCACCCG ACTTTGAGGG CGGGCGAGC GTCAAGCTAT GGTTGAGAC 360
 CCTGAGCGGC ATGTCGGCGT CAGTGAGCT GGACGACTCA CAGGTGGTTC AGATCTGTT 420
 10 CGACCTGGAG TCAGCCTACA ACGCCCTCAA CGCCCTCTG CATGCCCTAG CGGGGGCAC 480
 TAGCCCTTGC ACAGAAGGGC AGAGTCTGAG CGCATGGCTC CTGGTCTTCT GTGGGCCACA 540
 15 CAGGCCGTGG TCATCCACAC AACTCACTGT CTGAGCTGC CTGTCCTTGT TCTGCTTTG GTGTCAGAAC TTTTGGCCG GGCCCCCTCCC GCAATAAAG ATGCTCTTGG ACCTTCAAA 600
 AAAAAAA AAAAATCTRG GGGGGGCCCCG GTCCCAATCC CCCCNREI 660
 20

(2) INFORMATION FOR SEQ ID NO: 224:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1384 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

GGGAACTGC AGTGACAGCA GGAGTAAGAG TGGGAGGGAG GACAGAGTTG GGACACAGGT 60
 35 ATGGAGAGGG GGTTCAAGCGA GCCTAGAGAG CGCAGACTAT CAGGGTGGCG GCGGTGAGAA 120
 TCCAGGGAGA GGAGCGGAAA CAGAGAGGG CGAGAGGACC GGCCCCATTG TGGGTGAGAG 180
 40 AGCCCCCTAG CCATGTTGGG AGCCRAGCCA CACTGGCTAC CAGGTCCTT ACACAGTCAC
 GGGCTGCCCT TGGTTCTGGT GCTTCTGGCC CTGGGGGCGG GGTGGGCTCA GGAGGGTCA 300
 GAGCCCGTCC TCTTGGAGGG GGAGTGCCTG GTGGTCTGTG AGGCTGGCG AGTTCCTGCA 360
 45 GGGGGGGCCCG GGGGAGCAGC CCTGGAGAG AGACCCCTG CGCGAGTGGC ATTTCTGCG 420
 GTCCGAAGCC AMCACCATGA GCGCCAGGG GAAACCGCA ATGCCCTCAK TGGGGCCATC 480
 50 TACTTCGACC AGGTCTGGT GAACGAGGGC GGTGGCTTIG ACAGGGGCTC TGGCTCTTC
 GTAGCCCCCTG TCCGGGTGT CTACAGCTTC CGGTCCATG TGGTCAAGGT GTCACAAACGGC 540
 CAAACTGTCC AGGTGAGCCT GATGCTGAAC ACGTGGCTG TCATCTCAGC CTTTGGCAAT
 55 GATCCTGACG TGACCCGGGA GGCAGCCACC AGCTCTGTGC TACTGCCCTT GGACCCCTGGG 720
 GACCGAGTGT CTCTGCGCT CGCTCGGGGG ATCTCTCTGG GTGGTCTGAA ATGATCAAGT
 60 TTCTCTGGCT TCCTCATCTT CCCTCTCTGA GGACCCAAAGT YTTTCAAGCA CGAGAATCCA 780
 840

	GCCCCGACA ACTTTCTTCT GCCCCTCTT GCCCCAGAAA CAGCAGAGGC AGGAGAGAGA	900
	CTCCCTCTGG YTCCCTATCCC ACYTCTTGC ATGGGAMCCT GTGCCAAACA CCCAAGTTA	960
5	AGARAARARY ARARCTGWGG CAGGTATACA GAGCTGGAAG TGGACCATGG AAAACATSGA	1020
	TAACCATGCA TCYTCTTGCT TGGCACCTC CTGAAACTGT CCACCTTGA AGTTTGAACT	1080
	TTAGTCCCTC CAMACTCTGA CTGCTGCCCTC CTTCCTCCCA GCTCTCTCAC TGAGTTATYT	1140
10	TCACTGTACC TGTTCCAGCA TATCCCACAT ATCTCTCTTT CTCCCTGATCT GTGCTGTCIT	1200
	ATTCTCCTCC TTAGGCTTCC TATTACCTGG GATTCCATGA TTCATTCTT CAGACCCCT	1260
15	CCTGCCAGTA TGCTAAACCC TCCCTCTCTC TPTCTTATCC CGCTGTCCCA TTGGCCCAAGC	1320
	CTGGATGAAT CTATCAATAA AACAACTAGA GAATGGTGGT CAAAAAAA AAAAAAAAAC	1380
20	TCGA	1384

(2) INFORMATION FOR SEQ ID NO: 225:

25

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 760 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

35	GGGTGACCC ACGCGTCCGC TGACCAGTCC GTTATAGATA CTTCTTCCTA TACCAAAACT	60
	GTTTAAACAG GTGCCACCAAC AAGGGATGTC GTCCCTACTC TCTGCCGGTC TTCAAGCATC	120
	CCTTGTGGG AAARGTCTCT GGGCAAGCAC GTGGTATTG GTCTGCTGCT TGCTTCCCTT	180
40	TTTCCACCAAG GGATGTGIG ATCATAAGTC AAAACAACAG TATATTCCA ATCTCAAAG	240
	CTATTGTGGC CTGAGCACAA TTGAAATCTA GCAGAGTTTT TCCTATGTAG CTTTAGAGTA	300
	ACTCTTCTGC TTCTCTGICA CTTACAATTC AGGTTCTGCC TTTGCCTAAG AGCATGAGCA	360
45	GAAGAGTCCT CATGTGACGC TTAGTTCTAT TGCAGTCCTG GGTGAAACTA TTTAAGCWAT	420
	GGGGCTGCTK CTCCCCANWT CCTCCCTAAC AATTGTTGT GTGGACTTCT CATCTAAAAG	480
50	GTTAGTGGCT TTTGCTTGGG ATCAGTGCTC TCTATTGATG TTCTTGCCTG TCTCCAGACA	540
	CATTCTGTT GCATTAAGAC TTGAAAGACT TGTAGATGTG TGATGTCAG GCACAGGATG	600
	CTGAAAGCTA TGTTACTATT CTTAGTTGT AAATGTCCT TTTGATACCA TCATCTTGT	660
55	TTCTTTTGT AGGTATAAAAT AAAACACTG TTGACAATAA AAAAAAAA AAAAAAAA	720
	AAAAAAAAA AAAAAAAA NAAAAAAA AAAAAAAA	760

(2) INFORMATION FOR SEQ ID NO: 226:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2057 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

CCGAGCCGGC	TGCGCCGGGG	GAATCCGTGC	GGGOGCCTTC	CGTCCCRGTC	CCATCCTCGC	60
CGCGCTCCAG	CACCTCTGAA	GTTTGCGAGC	GCCCAGAAAG	GAGGCGAGGA	AGGAGGGAGT	120
GTGTGAGAGG	AGGGAGCAAA	AAGCTCACCC	TAAAACATT	ATTTCAAGGA	GAAAAGAAAA	180
AGGGGGGGCG	AAAAAATGGC	TGGGCAATT	ATAGAAAACA	TGAGCACCAA	GAAGCTGTGC	240
20 ATTGTTGGTG	GGATTCTGCT	CGTGTTCATA	ATCATCGCCT	TTCTGGTGGG	AGGCTTGATT	300
GCTCCAGGGC	CCACAACGGC	AGTGTCTAC	ATGTCGGTGA	AATGTGTGGA	TGCCCGTAAG	360
25 AACCATCACA	AGACAAAATG	GTTCGTGCCT	TGGGGACCCA	ATCATTGTGA	CAAGATCCGA	420
GACATTGAAG	AGGCAATTCC	AAGGGAAATT	GAAGCCAATG	ACATCGTGT	TTCTGTTTCAC	480
30 ATTCCCCCTCC	CCACATGGA	GATGAGTCCT	TGGTTCCAAT	TCATGMI GTT	TATCCTGCAG	540
CTGGACATTG	CCTTCAAGCT	AAACAACCAA	ATCAGRGAAGA	ATGCAGAAGT	CTCCATGGAC	600
GTTTCCCTGG	CITACCGTGA	TGACCGGTTT	GCTGAGTGGG	CTGAAATGCC	CCATGAAAGA	660
35 GTACCACGGA	AACTCAAATG	CACCTTCACA	TCTCCCAAGA	CTCCAGAGCA	TGGAGGGCCG	720
GTTACTATGA	ATGTGATGTC	CTTCCCTTC	TGGAAATTGG	GTCTGTGCC	CATGAAGTTT	780
40 TACCTTTAA	ACATCCGGCT	GCCTGTGAAT	GAGAAGAAGA	AAATCAATGT	GGGAATTGGG	840
GAGATAAAGG	ATATCCGGTT	GGTGGGGATC	CACCAAAATG	GAGGCTTCAC	CAAGGTGTGG	900
TTTGCATGA	AGACCTTCCT	TACGCCAGC	ATCTTCATCA	TTATGGTGTG	GTATTGGAGG	960
45 AGGATCACCA	TGATGTCCCC	ACCCCCAGTG	CTTCTGGAAA	AAGTCATCTT	TGCCCTTGGG	1020
ATTTCATGA	CCTTTATCAA	TATCCCAGTG	GAATGGTTTT	CCATCGGGTT	TGACTGGACC	1080
50 TGGATGCTGC	TGTTGGTGA	CATCCGACAG	GCATCTCTA	TGCRATGCTT	CTKTCCTTCT	1140
GGATCATCTT	CTGTGGCGAG	CACATGATGG	ATCAGCACGA	GCGGAACAC	ATCGCAGGGT	1200
ATTGGAAGCA	AGTCGGACCC	ATTGCCGTG	GTCCTCTG	CTCTTCATAT	TTGACATGTG	1260
55 TGAGAGAGGG	GTACAACCTCA	CGAATCCCTT	CTACAGTATC	TGGACTACAG	ACATTGGGAA	1320
CAGAGCTGGC	CATGGCTTTC	ATCATCGTGG	CTGGAATCTG	CCTCTGCCTC	TAACCTCCTG	1380
60 TTTCTATGCT	TCATGGTATT	TCAGGTGTTT	CGGAACATCA	GTGGGAAGCA	GTCCAGCCTG	1440

	CCAGCTATGA GCAAAGTCCC CCGGCTACAC TATGAGGGGC TAATTTCAG GTTCAAGTTC	1500
	CTCGAGCTTA TCAACCTTGGC CTGGCGTCCC ATGAGTGTCA TCTTCTTCAT CGTTAGTCAG	1560
5	GTAACGGAGG CCCTTTGGCA ATGGGGGCC CGTTCAGTC CCAAGTGAAC AGTGCCMTT	1620
	TCTAAGGCTAT CTATGGGATG TCGAAATCTGT ATGCTTTCGC TCTGATGTT TCATGATGCAC	1680
10	CATCCCTAA AAACATATGA GAGAACCTT CCTTCTGGAT GCAACTCCCC TGAAATTCGA	1740
	CGGAGMTG TGGTTGGTT GTTTCGGHAC TTATCAGCA ATGTTTCAGC GCTTCGAAAT	1800
	ATTCCTTCAT CAATGAGAAC CGAGCTTCG GATTTGGT CACAAGGCA ACACATGTTT	1860
15	ATCGCTTTCG CTTCGGAGT TCTCACAGTC AGCTTGGTG TACTTGTATA CCCACACAAA	1920
	TACACTCACT TACGCTTAT CTGAAATGT TAATATAGA GAAAAAAGCG TCAACAATAA	1980
20	ATATTCCTTG ACTATGTGT TACTTCCTT AAAAAAAAAA AAAAAAAACTC GTGCCGAATT	2040
	CGCCCGGAGC GGCAAGCA	2057

25

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2084 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

35	GGCAGAGGGC CATTCTGGC AAGAGGCCAA ACCCCCCATTC CTCTGTGCCCT CTCCCTCTCCC	60
	ACCAAGTGT TTTTAAAT ATGCTTGTGTT ACCGGAAATA ACTGTTCAATT TTTCACCTCCT	120
40	CCCTCCTAGG TCAACTTTT CAGAAAGAA ATCTGCCTCC TGGAAACCAG AAGAAAATA	180
	TGAGACGGGG AATCTCGTC TGTGTGTGT SCTGCCCTTG GCTGAGTGTG TGGAGTCCTG	240
45	CTCAAGGTGT AGGTACAGTG TGTGTGATCG TGGGGCTTG AGGGAAACCG CTGTTTCAGA	300
	GCTGTGACTG CGCTCTCACT GCAAGAGAASC TGCCCTTGGC TGCTCGTAGC GCGGGGCCTT	360
	CTCTCCTCTGT CTCATCGAG ACCAGCCAGT GTCCGGGGGG CAGAAGGTAC CGGGGCAGCT	420
50	ACTGGAGGGAC TGTGGGGGAC TGCCTGGCT GCCCCCTCTCG CGGTGGGGCC CTGTTGCTGC	480
	TGTCCTCTCA TTCTACTAC TCCCTCCAA ATGCGGTCTGG CCCGCCCTTC ACTTGGATGC	540
	TTGCCCTCTCT GGGCTCTTC GCAAGGCACTG AACATCCCTCC TGGGCCCAA GGGCCTGGCC	600
55	CCAGCTGACA TCTCTGCACT GTGTGAATTA GGGATTTCAC CGTGGGCCAA TGGGCTGGCA	660
	TGGTCATATT ACATCGGACA TGTGGGGCTG ATCCTGCCAG AGCTCCAGGC CGGGATTCTGA	720
60	ACTTACAAATC AGCAATTACA CAACTGCTA CGGGGTGCCAG TGAGCCAGCG GTGTNATATT	780

	CCTCTCCAT TGGACTGTGG GGTGCCTGAT AACCTGAGTA TGGCTGACCC CAACATTGCG	840
5	TTCCTGGATA AACTGCCCA GCAGACCGGT GACCGTGCTG GCATCAAGGA TCGGGTTTAC	900
	AGCAAACAGCA TCTATGAGCT TCTGGAGAAC GGGCAGCGGG CGGGCACCTG TGTCTGGAG	960
	TACGCCACCC CCTTGGAGAC TTTGGTTGCC ATGTCACAAT ACAGTCAAGC TGGCTTTAGC	1020
10	GGGGAGGATA GGCTTGACCA GGCCAAACTC TTCTGCCGG AACTTGAGGA CATCCTGGCA	1080
	GATGCCCTG AGTCTCAGAA CAACTGCCGC CTCATTGCC ACCAGGAACC TGCAGATGAC	1140
15	AGCAGCTCT CGCTGTCCA GGAGGTTCTC CGGCACCTGC GGCAGGGAGGA AAAGGAAGAG	1200
	GTTACTGTGG GCAGCTTGAA GACCTCAGCG GTGCCAGTA CCTCCACGAT GTCCAAGAG	1260
	CCTGAGCTCC TCATCAGTGG AATGGAAAAG CCCCTCCCTC TCCGCACGGG TTTCTCTGA	1320
20	GACCCAGGGT CACCAGGCCA GAGCCTCCAG TGGCTCCAA GCCTCTGGAC TGGGGGCTCT	1380
	CTTCAGTGGC TGAATGTCCA GCAGAGCTAT TTCCCTCCAC AGGGGGCTT GCAGGGAAAGG	1440
25	GTCCAGGACT TGACATCTTA AGATGCGTCT TGTCCCCCTG GGCCAGTCAT TTCCCCCTCTC	1500
	TGAGCCTCGG TGTCTTCAAC CTGTGAAATG GGATCATAAT CACTGCCCTA CCTCCCTCAC	1560
	GGTGTGTTGTG AGGACTGAGT GTGTGGAAGT TTTTCATAAA CTTTGGATGC TAGTGTACTT	1620
30	AGGGGGTGTG CCAGGTGTCT TTCATGGGGC CTTCCAGACC CACTCCCCAC CCTTCTCCCC	1680
	TTCCCTTGGCC CGGGGACGCC GAACCTCTC AAATGGTATCA ACAGGCTCT TCGCCCTCTG	1740
35	GCTCCTGGTC ATGTTCCATT ATTGGGGAGC CCCAGCAGAA GAATGGAGAG GAGGAGGAGG	1800
	CTGAGTTGG GGTATTGAAT CCCCCGGCTC CCACCCCTGCA GCATCAAGGT TGCTATGGAC	1860
	TCTCCTGCCG GGCAACTCTT GCGTAATCAT GACTATCTCT AGGATTCTGG CACCACTTCC	1920
40	TTCCCTGGCC CCTTAAGCCT AGCTGTGTAT OGGCACCCCC ACCCACTAG AGTACTCCCT	1980
	CTCACCTGGCG GTTTCTTAT ACTCCACCCC TTTCTCAACG GTCTTTTTT AAAGCACATC	2040
45	TCAGATTAAA AAAAAAAA AAAAAAAA AGGGGGGGCN GCNT	2084

(2) INFORMATION FOR SEQ ID NO: 228:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2143 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

60

TCGACCCACG CGTCCGGGTG AATTCCTTGA CCTGCAAACA CATATTTATT AGCCTGACTC

60

	AAACAATGAA GCTATTAAAA CTTCGGAGGA ACATTGTAAA ACTCTCTTG TATCGGCATT	120
	TCACCAACAC GCTTATTTG GCAGTGGCAG CATCCATTGT GTTTATCATC TGGACAACCA	180
5	TGAAGTCAG AATAGTGACA TGTCAAGTCGG ACTGGCGGA GCTGTGGTA GACGATGCCA	240
	TCTGCGCTT CCTGTTCTCC ATGATCCTCT TTGTCATCAT GGTTCTCTGG CGACCACCTG	300
10	CAAACAACCA GAGGTTGCC TTTTACCAT TGTCTGAGGA AGAGGAGGAG GATGAACAAA	360
	AGGAGCCTAT GCTGAAAGAA AGCTTGAAG GAATGAAAAT GAGAAGTACC AAACAAGAAC	420
	CCAATGGAAA TAGTAAAGTT AACAAAGCAC AGGAAGATGA TTTGAAGTGG GTAGAAGAGA	480
15	ATGTCCTTC TTCTGTGACA GATGTAGCAC TTCCAGCCCT TCTGGATTCA GATGAGGAAC	540
	GAATGATCAC ACACTTGAA AGGTCCAAA TGGAGTAAGG AATGGGAAGA TTTGCAGTTA	600
20	AAGATGGCTA CCATCAGGGAGAGATCAGC ATCTGIGTCA GTCTTCTGTA CGGCTCCATG	660
	GGATTAAGG AAGCAATGAC ATCCTGATCT GTTCCTTGAT CTTTGGCAT TGGAGTTGGC	720
	GAGAGGTGTC AGAACAAAGA GAACATCTTA CTGAAAACAA GTTCATAAGA TGAGAAAAT	780
25	CTACGAGCTT CTTATTTACA ACACGTGTC CCCCTTCCT CCCAGACTCT GACATGGATG	840
	TTCATGCAAC TTAAGTGTGT TGTTCTGAA CTTTCTGAA TGTTCAATT TTTAAATCTG	900
	ACAAACTAAA AAGTTAACG TCTTCTAAA GATTGTACATC AACACCATAA TATGTAATCT	960
30	CCAGGAGCAA CTGCCTGAA TTTTATTTA TTTAGGGAGT TACATAGGTG ATGGGGAAA	1020
	TTGTTAACTA CCTTTCATTT TCCTGGGAAG TCAAGGTTAC ATCTTGCAGA GGTGTTTGT	1080
35	AGAAAAAAGG GCCCTCTGA GTTAAGGAGC CATAGTCTA TCAATGATCA AAAGAAAAAA	1140
	AAAAAAAAGA GAAACTGTTA CAGTATGATT CAGATCATT AAAAAGCAA AATCAAGTGC	1200
	ATTTTGTAAACAAATGGTG TATATTAAG ATTTTCTAT TTCAAGATGTA CTTAAAGAG	1260
40	AAATATTAGC TTAACTCTTT TGACATCTGC TATTGTGACA CATCCCATTG CTGGCAATGT	1320
	GGTGCACACT CCGAAACTTT TAACTACTGT TTTGTAAGCC TCCAAGGGTG GCATTGCAGG	1380
45	GTCCCTAGGC AATGTTTGT TTGCTTTAT GCAGAGAGGT GCTCCAAGTG CTGTGATTGA	1440
	GCACCGTGCT AGAGGAACGT TAATGCTCA GAAGTTGTAG CTTATACAAA GGAAACAGGT	1500
	CCTGCTGGCT TAATTTAAC AGTTATTGCA TGAAGTACCG TGGAGGCCCT GGACTGCTGC	1560
50	TCGTTCTTA GGATGGACTG TTCTGGTATC TGGTATTGGT TTAGAGACTG TTAATAAGGG	1620
	ACATCACAAG GTGATGGGAT TCATTGAAAG CACTCTATT CTGTTTAAT GGTTTATCC	1680
55	AATTTGCCT TCCCAAGATT TTTGTTCTAC ATAAAAAGTT CATGCCACTT TTTAATATAA	1740
	AAAAATTAA CAAAATTAAT GTATTTCTT CATTTCCTTAAACTTTTCA TAAAGACTCT	1800
	TTCTGTCAAA CTCATGAAAA ATTCTTTCTT ATGGCTTTA TTCTAGATTG TCTTATTTTC	1860
60		

TGTTAAAACC AATGACCACA TGACCACAAT CTTCACTAAC TCATACTGCA GTGAAAGTGT	1920
TAACCCCTAG GTAGTTCTC TACAACCTCT TGCTATGGTG ATTTTTAAAAA AAGTTTCCTA	1980
5 GGGAAAGTATC TCTGAGGGAA CAGGCAATCT GAAGGAAC TG ACTATATTCT CCATGGCTAA	2040
GTCCATTAGG CAAAAGNCT GGGTGGGTAT TGTTTGTCA GCTGTCTATT GGCATAATTAA	2100
AAACGTAGGC CGGANGGAAT AAATAGGTG TNATGCCGGC GGG	2143
10	

10

(2) INFORMATION FOR SEQ ID NO: 229:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1025 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

CCTGGCCCAC ATTGCTTCAT TGGCCTGGCC ATGCGCCTGT ACTATGGCAG CCCCTAGTCC	60
25 CTGACAAC TT CCACCCCTGAT TCCGGACCCCT GTAGATTGGG CGCCACCCACC AGATCCCCCT	120
CCCAGGGCTT CCTCCCTCTC CCATCAGCAG CCCTGTAAACA AGTGCCTTGT GAGAAAAGCT	180
30 GGAGAAAGTGA GGGCAGCCAG GTTATTCTCT GGAGGTTGGT GGATGAAGGG GTACCCCTAGG	240
AGATGTGAAG TG TGGGTTTG GTTAAGGAAA TGCTTACCAT CCCCCACCCC CAACCAAGTT	300
35 CTTCCAGACT AAAGAATTAA GGTAACATCA ATACCTAGGC CTGAGAAATA ACCCCATCCT	360
TGTTGGGCAG CTCCCTGCTT TGTCCCTGCAT GAACAGAGTT GATGAAAGTG GGGTGTGGGC	420
AACAAGTGGC TTTCCTTGCC TACTTTAGTC ACCCAGCAGA GCCACTGGAG CTGGCTAGTC	480
40 CAGCCCAGCC ATGGTGCATG ACTCTTCCAT AAGGGATCCT CACCCCTCCA CTTTCATGCA	540
AGAAGGCCCA GTTGCACAG ATTATACAAC CATTACCCAA ACCACTCTGA CAGTCTCCTC	600
45 CAGTTCCAGC AATGCCTAGA GACATGCTCC CTGCCCTCTC CACAGTGCTG CTCCCCACAC	660
CTAGCCCTTIG TTCTGGAAAC CCCAGAGAGG GCTGGGCTTG ACTCATCTCA GGGAAATGTAG	720
CCCCCTGGGCC CTGGCTTAAG CCGACACTCC TGACCTCTCT GTTCACCCCTG AGGGCTGTCT	780
50 TGAAGCCCGC TACCCACTCT GAGGCTCTA GGAGGTACCA TGCTTCCCAC TCTGGGGCCT	840
GCCCCCTGCCT AGCAGTCTCC CAGCTCCAA CAGCCTGGGG AAGCTCTGCA CAGAGTGACC	900
55 TGAGACCAGG TACAGGAAAC CTGTAGCTCA ATCAGTGTCT CTTTAACCTGC ATAAGCAATA	960
AGATCTTAAT AAAGTCTCT AGGCTGTAGG GTGGTTCCCTA CAACCCACAGC CAAAAAAA	1020
AAAAA	1025

60

(2) INFORMATION FOR SEQ ID NO: 230:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1250 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

GCCACGGGT CGGCCCACGC GTCGGCGGT GCGGAGTATG GGGCGCTGAT GGCCATGGAG	60
GGCTACTGGC GCTTCCTGGC GCYGCTGGGG TCGGCACITGC TCGTCGGCTT CCTGTCGGTG	120
ATSTTCGCC CCGTCTGGGT CCTCCACTAC CGAGAGGGGC TTGGCTGGGA TGGGAGCGCA	180
CTAGAGTTA ACTGGCACCC AGTGCTSATG GTCACCGGCT TCGTCITCAT CCAGGGCATC	240
GCATCATCGT CTACAGACTG CCGTGGACCT GGAAATGCAG CAAGCTCCTG ATGAAATCCA	300
TCCATGCAGG GTTAAATGCA GTTGCTGCCA TTCITGCAAT TATCTCTGTG GTGGCCGTGT	360
TTGAGAACCA CAATGTTAAC AATATAGCCA ATATGTACAG TCTGCACAGC TGGGTTGGAC	420
TGATAGCTGT CATATGCTAT TTGTTACAGC TTCTTCAGG TTTTCAGTC TTCTGCTTC	480
CATGGGCTCC GCTTTCTCTC CGAGCATTTC TCATGCCCAT ACATGTTAT TCTGGAATTG	540
TCATCTTGG AACAGTGATT GCAACAGCAC TTATGGGATT GACAGAGAAA CTGATTTTT	600
CCCTGAGAGA TCCTGCATAC AGTACATTCC CGCCAGAAGG TGTTTCGTA AATACGCTTG	660
GCCTTCTGAT CCTGGTGTTC GGGGCCCTCA TTTTTGGAT AGTCACCAGA CCGCAATGGA	720
AACGTCTAA GGAGCCAAAT TCTACCATTTC TTCATCCAAA TGGAGGCAGT GAACAGGGAG	780
CAAGAGGTTTC CATGCCAGCC TACTCTGGCA ACAACATGGA CAAATCAGAT TCAGAGTTAA	840
ACATGAACT AGCAGCAAGG AAAAGAAACT TAGCTCTGGA TGAGGCTGGG CAGAGATCTA	900
CCATGTAAAA TGTTGTAGAG ATAGACCAT ATAACGTAC GTTTCAAAAC TAGCTCTACA	960
GTTTTGCTTC TCCTATTAGC CATATGATAA TTGGCTATG TAGTATCAAT ATTACTTTA	1020
ATCACAAAGG ATGGTTCTT GAAATAATT GTATTGATTG AGGCCTATGA ACTGACCTGA	1080
ATTGGAAAGG ATGTGATTAA TATAAATAAT AGCAGATATA AATTGIGGTT ATGTTACCTT	1140
TATCTTGTG AGGACCACAA CATTAGCACG GTGCCCTGTG CAKAATAGAT ACTCAATATG	1200
TGAATATGTG TCTACTAGTA GTTAATTGGA TAAACTGGCA GCATCCCTGA	1250

55

(2) INFORMATION FOR SEQ ID NO: 231:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1811 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

CNGNCAGTAC CGGTCNGATT CCCGGGTGGA CCCACGCGTC CGCTGCATTTC	60
10 CAGTGGCTTT CATTCTGAAG TTCCCTGGATA ACATGTTCCA TGTCTTGATG GCCCAGGGTA	120
CCASTGTCAT TATCACAACA GTGTCGTGCC TGGTCTTTGA CTTCAGGCC TCCCTGGAAT	180
15 TTTCTTGGAA AGCCSCATCA GTCSTYCTCT CTATATTTAT TTATAATGCC AGCAAGCCTC	240
AAGTCCGGA ATACGCCACCT AGGCAAGAAA GGATCCGAGA TCTAAGTGGC AATCTTTGGG	300
ACCGTTCCAG TGGGGATGGA GAAGAACTAG AAAGACTTAC CAAACCCAAG AGTGATGAGT	360
20 CAGATGAAGA TACTTTCTAA CTGGTACCCA CATAGTTGC AGCTCTCTTG AACCTTATTT	420
TCACATTTTC AGTGTGTTGTA ATATTTATCT TTTCACTTIG ATAAACCAGA AATGTTTCTA	480
25 AACCTTAATA TTCTTTGCAT ATATCTAGCT ACTCCCTAAA TGTTCCATC CAAGGCTTAG	540
AGTACCCAAA GGCTAAGAAA TTCTAAAGAA CTGATACAGG AGTAACAATA TGAAGAAC	600
ATTAATATCT CAGTACTTGA TAAATCAGAA AGTTATATGT GCAGATTATT TTCCCTGGCC	660
30 TTCAAGCTTC CAAAAAACTT GTAATAATCA TGTTAGCTAT AGCTTGTATA TACACATAGA	720
GATCAATTG CCAAATATTIC ACAATCATGT AGTTCTAGTT TACATGCCAA AGTCTTCCCT	780
35 TTTAACATT ATAAAAGCTA GGTTGTCTCT TGAATTTGA GCCCCTAGAG ATAGTCATT	840
TGCAAGTAAA GAGCAACGGG ACCCTTTCTA AAAACGTTGG TTGAAGGACC TAAATACCTG	900
GCCATACCAT AGATTTGGGA TGAITGTAGTC TGTGCTAAAT ATTTGCTGA AGAAGCAGTT	960
40 TCTCAGACAC AACATCTCAG AATTTTAATT TTTAGAAATT CATGGAAAT TGGATTTTG	1020
TAATAATCTT TTGATGTTT AACATTTGGT TCCCTAGTCA CCATAGTTAC CACTTGTATT	1080
45 TTAAGTCATT TAAACAAGCC ACGGTGGGGC TTTTTCTCC TCAGTTGAG GAGAAAATC	1140
TTGATGTCAT TACTCCTGAA TTATTACATT TTGGAGAATA AGAGGGCATT TTATTTTATT	1200
AGTTACTAAT TCAAGCTGTG ACTATTGTAT ATCTTTCCAA GAGTTGAAAT GCTGGCTTCA	1260
50 GAATCATAACC AGATTGTCAG TGAAGCTGAT GCCTAGGAAC TTTAAAGGG ATCCCTTCAA	1320
AAGGATCACT TAGCAAACAC ATGTTGACTT TTAACTGATG TATGAATATT AATACTCTAA	1380
55 AAATAGAAAG ACCAGTAATA TATAAGTCAC TTTACAGTGC TACTTCACAC TTAAAAGTGC	1440
ATGGTATTTT TCATGGTATT TTGCTATGCCAG CCAGTTAACT CTCGTAGATA GAGAAGTCAG	1500
GTGATAGATG ATATTAAGAA TTAGCAAACA AAAGTGACTT GCTCAGGGTC ATGCAGCTGG	1560
60 GTGATGATAG AAGAGTGGGC TTTAAGTGGC AGGCCTGTAT GTTACAGAC TACCATACTG	1620

TAAATATGAG CTTTATGGTG TCATTCTAG AAACCTATAAC ATTTCTGCTC TCCTTTCTCC	1680
TAAGTTTCAT GCAGATGAAT ATAAGGTAAT ATACTATTAT ATAATTCATT TGTGATATCC	1740
5 ACAATAATAT GACTGGCAAG AATGGTGGA AATTGTAAT TAAAATAATT ATTAAACCTA	1800
AAAAAAAAN N	1811

10

(2) INFORMATION FOR SEQ ID NO: 232:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2271 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CTGACCTCAT GCGTAGAGC CTAGCAACAG CGCAGGCTCC CAGCCGAGTC CGTTATGGCC	60
25 GCTGCCGTCC CGAACAGGAT GAGGGGGCCA GCACAAGCGA AACTGCTGCC CGGGTCGGCC	120
ATCCAAGCCC TTGTGGGTT GGCGGGGGCG CTGGTCTTGG CGCTCTGCT TGTGTCCGCC	180
30 GCTCTATCCA GTGTTGTATC ACGGACTGAT TCACCGAGCC CAACCGTACT CAACTCACAT	240
ATTCTACCC CAAATGTGAA TGCTTTAACCA CATGAAAACC AAACCAAACC TTCTATTTCC	300
CAAATCAGCA CCACCCCTCCC TCCCACGACG AGTACCAAGA AAAGTGGAGG AGCATCTGTG	360
35 GTCCCTCATC CCTCGCCCTAC TCCCTCTGCTC CAAGAGGAAG CTGATAACAA TGAAGATCCT	420
AGTATAGAGG AGGAGGATCT TCTGATGCTG AACAGTTCTC CATCCACAGC CAAAGACACT	480
40 CTAGACAATG GCGATTATGG AGAACCAAGAC TATGACTGGA CCACGGGGCC CAGGGACGAC	540
GACGAGTCIG ATNGACACCT TGGAAAGAAAA CAGGGGTTAC ATGGAAATTG AACAGTCAGT	600
GAAATCTTTT AAGATGCCAT CCTCAAATAT AGAAGAGGAA GACAGCCATT TCTTTTTICA	660
45 TCTTATTATT TTTGCTTTT GCATGCTGT TGTGTTACATT ACATATCACA ACAAAAGGAA	720
GATTTTTCTT CTGGTTCAA GCAGGAAATG GCGTGATGGC CTTTGTCCA AAACAGTGGA	780
50 ATACCATCGC CTAGATCAGA ATGTTAATGA GGCAATGCCT TCTTGAAGA TTACCAATGA	840
TTATATTTTT TAAAGCACTG TGATTTGAAT TTGCTTATGT AATTTTATTT GCTTGACTTT	900
TTATATGATA TTGIGCAAAT GTTGCCATA GGCAATTGGT ACTTAAATGA GAGGTGAGTC	960
55 TCTCTTTGCG CTTGGIGCTT TGGAAATTAA ATGTCACAAA CGAGTATATA ATTTTTTATC	1020
TGTACTTTTA GAGCTGAGTT TAATCAGGTG TCCAAAATGT GAGTTAAACA TTACCTTATA	1080
60 TTACACTGT TAGTTTTAT TGTGTTAGAT TTATATGCT TCTTCTGGAA GTATTAGTGA	1140

	TGCTACTTTT AAAAGATCCC AAACITGTAA CTAAATTCTG ACATATCTGT TACTGCTGAC	1200
	TCACATTICAT TCTCCGCCAT TCAAATACTA TTTTTTATCC ACATTTTTTT TTGTTCCCAA	1260
5	ACTGTAATGT ACAAGGATAT GTGTGATAAT GCTTTGGATT TGAGTAATAT TTTTTTTCT	1320
	TCCAAGAAAA CTGCTTTGGA TATTTTTAGA TAATTTAAC ATAATTAGG ATAATGATAT	1380
10	TGCTCAATCT GACCACAATT TTAGGTAAAA CATTAAATGT GTCAAGAAAT CTTGGCAACA	1440
	GAGACTCTGC AGCTTGCAGT GGACATAGAT AAAATGTAC AGAGATACTA TTTTTTTGGT	1500
	TGGAATTACT ATATTAATT TAGAACCAGA AACTGGTAAA ATGTTAAATA CATGTACAAT	1560
15	TGCTTTAGT TAGCAATTGA TTGTAGCATG GGTCCTCCA AGGTTCAAG CAATGGCAG	1620
	AGTTTAAAT TATATCAGAT TCGTTTACTT CGTTTATTAT TTTACAGTAA ATTGAAATAA	1680
20	ATCTTAGGGG TCATTATCAC TAAATAATA CTGTACCTAG GTCTTCAAA TTAAATTAT	1740
	ACCTGAATGA AGTTGTTGT ATACATAAAG GATATTGTG TACAATTACC TTTTTTCCCC	1800
	CACACTTGTGTT TTCTTTGTT TTGTTTTTA TGGCAACTGG AAAGTATTAA CTATGGGATT	1860
25	CATTATGTC TGTCTTCTA TCATAAAGAA TTGATCAATA TGAAATATG TGATTTGAAC	1920
	CATGGTGAC TTACAAGTGT CACTACAGCT TTTTAGAAAA CATAGCCCTA ATATATGTTA	1980
30	AGCAGGACCC GGGTGAGCCA GTGGGCTTGC GCTTTATGTA GAGCTGGAAG AAGGCCGTCC	2040
	ATCCTGTCTC TTGGGCGGAC AGTGTACTTT CCTAATAGGG AAGGGAGCA CAATGGAAAT	2100
	ACCCCTGAAC CGTTTTATTG CAGTAATTAA TTTCATATCT GAAACTATTA TTTAATATT	2160
35	TGAATAAGAT TTAAAAAAAT AAATGGCAAA GATATAAAC TAAAAAAAAA AAAAAAAAAA	2220
	AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAANANA N	2271

40

(2) INFORMATION FOR SEQ ID NO: 233:

45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1338 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:	
	CTTCGGTTC TCCGGGCAGC TGCCACTGCT GTAGCTTCTG CCACCTGCCA CGACCGGGCC	60
	TCTCOCTGGC GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCCCTCTTGGA	120
55	GCCAGCGTGG CGNGCCTGGC GGCTCCCGGG TCGTGAGAGA GCGGTCCGGG AACGATGAAAG	180
	GCCTCGCAGT GCTGCTGCTG TCTCAGGCCAC CTCTTGGCTT CCGTCCTCCT CCTGCTGTTG	240
60	CTGCCTGAAC TAAGCGGGYC CCTGGMAGTC CTGCTGCAGG CAGCCGAGGC CGGCCAGGT	300

	YTTGGGCCTC CTGACCCTAG ACCAGGACAT TACCGCCGCT GCCACCGGGC CCTWACCCCT	360
	GCCCCAGCAGC CGGGCCGTGG TCTGGCTGAA GCTGGGGGG CCGGGGGGCT CCGAGGGAGG	420
5	CAATGGCAGC AACCCCTGTGG CGGGGCTTGA GACGGACGAT CACGGACCGA AGGCCGGGA	480
	ARGCTCGGTG GGTGGCCGCC TTGCTGTGAG CCCCAACCCCT GGCGACAAGC CCATGACCCA	540
10	GCGGGCCCTG ACCGTGTTGA TGGTGGTGAG CGGCGCGGTG CTGGTGTACT TCGTGGTCAG	600
	GACGGTCAGG ATGAGAAGAA GAAACCGAAA GACTAGGAGA TATGGAGTTT TGACACTAA	660
	CATAGAAAAT ATGGAATTGA CACCTTTAGA ACAGGATGAT GAGGATGATG ACAACACGTT	720
15	GTTTGATGCC AATCATCCTC GAAGATAAGA ATGTGCCTTT TGATGAAAGA ACTTTATCTT	780
	TCTACAATGA AGAGTGGAAAT TTCTATGTTT AAGGAATAAG AAGCCACTAT ATCAATGTTG	840
20	GGGGGGTATT TAAGTTACAT ATATTNAAC AACCTTTAAT TTGCTGTTGC AATAAAATACC	900
	GTATCCTTTT ATTATATCCTT TATATGTATA GAAGTACTCT GTTAATGGGC TCAGAGATGT	960
	TGGGGATAAA GTATACTGTA ATAATTATAC TGTTTGAAAAA TTACTATAAA ACGGTGTTTT	1020
25	CTGRTCGGTT TTGTTTCCT GCTTACCAT A TGATTGTAAA TTGTTTTATG TATTAATCAG	1080
	TTAATGCTAA TTATTTTGC TGATGTCATA TGTTAAAGAG CTATAAATTC CAACAACCAA	1140
30	CTGGTGTGTA AAAATAATT AAAATYTCCCT TTACTGAAAG GTATTCCC A TTTTGTTGGG	1200
	GAAAAGAAC CAAATTATTACTTTGTTTGGGGTTTTA AAATATTAAG AAATGTCTAA	1260
	GTTATTGTTT GCAAAACAAT AAATATGATT TTAAATTCTC TTAAAAAAA AAAAAAAAC	1320
35	CCCCCCGGG GGGCCGGN	1338

40

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

Met	Leu	Ser	Thr	Gly	Ile	Glu	Val	Ala	Arg	Pro	Pro	Ala	Thr	Leu	Leu
1															

5							10								15
---	--	--	--	--	--	--	----	--	--	--	--	--	--	--	----

Gly	Leu	Met	Phe	Val	Leu	Thr	Gly	Met	Pro	Arg	Gly	Leu	Arg	Xaa
20														30

55

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 116 amino acids

60

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

5 Met Asn Val Val Ile Val Ile Ile Leu Phe Ser Phe Asp Ser Val Gly
1 5 10 15

10 Thr Met Phe Ser Cys Asn Arg Ile Pro Lys Ile Thr Val Leu Asn Lys
20 25 30

10 Leu Lys Phe Xaa Cys Glu Val Leu Leu Arg Ile Gln Thr Ile Gln Gly
35 40 45

15 Phe Tyr Arg Cys Thr Arg Ile Ser Arg Tyr Lys Gly Ile Phe Pro Asp
50 55 60

Phe Cys Gln Ser Gln Cys Met Gly Cys Asn Pro Glu Ser Xaa Met Ala
65 70 75 80

20 Val Pro Ala Leu Val Thr Pro Ile Leu Ala His Arg Lys Lys Glu Lys
85 90 95

Gly Met Cys Leu Phe Thr Leu Ile Ile Ala Pro Thr Arg Cys Thr His
100 105 110

25 Tyr Phe Cys Xaa
115

30 (2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

35 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

40 Met Ser Ser Ala Lys Ile Val Arg Gln Arg Gly Ala Val Pro Thr Tyr
1 5 10 15

Tyr Thr Thr Glu Ala Gly Glu Ile Ile Phe Leu Val Leu Asn Trp Ser
20 25 30

45 Leu Ser Ile Leu His Ile Val Asp Val Leu Cys Ser Lys Pro Glu Lys
35 40 45

Ser Val Thr Glu Asp Ala Ala Ser Gly Leu Ser Gln Arg Met Thr Ala
50 55 60

50 Leu Val Trp Arg Lys Gly Pro Asp Gly Gly Ser Arg Lys Pro Ile Leu
65 70 75 80

55 Leu Leu Phe Phe Leu Pro Leu Ile Leu Cys Phe His Ser Phe Ile
85 90 95

His Ser Ser Asn Ile Cys Xaa
100

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

10 Met Ile Leu Phe Pro Gln Xaa Ala Leu Arg Leu Gly Xaa Trp Pro Arg
1 5 10 15

Thr Trp Ser Ile Leu Xaa Lys Tyr Ser Val Asn Phe Phe Ser Ala Tyr

20 25 30

Ser Pro Met Gly Ala Val Gly Thr Glu Phe
35 40

20 (2) INFORMATION FOR SEO ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

Met Ile Ile Leu Leu Leu Phe Met Leu Leu Asn Asn Val Val Val Leu Val
20 1 5 10 15

Gln Glu Asp Asn Cys Gln Arg Lys Asn Thr Val Gln Glu Arg Arg Xaa
20 25 30

35 Trp Ser Gln Trp Xaa
35

40 (2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

Met Ala Ala Xaa Pro Pro Gly Cys Thr Pro Pro Xaa Leu Leu Asp Ile
1 5 10 15

50 1 5 10 15
Ser Trp Leu Thr Glu Ser Leu Gly Ala Gly Gln Pro Val Pro Val Glu

55 Cys Arg His Arg Leu Glu Val Ala Gly Pro Arg Lys Gly Pro Leu Ser
56 57 58 59 60 61 62 63 64 65

Pro Ala Trp Met Pro Ala Tyr Ala Cys Gln Arg Pro Thr Pro Leu Thr
50 55 60

60 His His Asn Thr Gly Leu Ser Glu Leu Leu Glu His Gly Val Cys Glu

490

65 70 75 80
Glu Val Glu Arg Val Arg Arg Ser Glu Arg Tyr Gln Thr Met Lys Val
85 90 95

5 Arg Arg Ala Gly Leu Gly Pro Thr Pro Gly Met Ser Cys Pro Gly Asn
100 105 110

10 Asp Asn Thr Val His Thr Met His Gly Glu Ala Asn Arg Gly Ser Xaa
115 120 125

15

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 67 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

25 Met Ser Ile Leu Cys Cys Pro Xaa Leu Cys Leu Phe Phe Ser Phe Cys
1 5 10 15

Ile Ser Ser Gly Ser Cys Pro Phe Ser His Val Ser Gln Leu Ser Phe
20 25 30

30 Ile Ala Thr Phe Ser Gln Ser Ser Pro Val Leu Leu Val Pro Ala Tyr
35 40 45

35 Asn Thr Tyr Leu Ser Phe Leu Ala Phe Leu Asp Cys Ala Ser Leu Thr
50 55 60

Ser Thr Xaa
65

40

(2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 69 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

50 Met Ser Thr Phe Gln Leu Leu Leu Ile Leu Ala Gln Ser Thr Tyr
1 5 10 15

Lys Ile Lys Ser Lys Pro Leu His Met Thr Asn His Thr Leu Leu Asn
20 25 30

55 Ser Pro Gly Leu Asn Pro Ser Ser Pro Thr Leu Asn Phe Lys Thr Gln
35 40 45

60 Gln His Glu Ser Val Ser Tyr Ala Cys Cys His Met Arg Ser Leu His
50 55 60

His Ala Phe Ala Xaa
65

5

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

15 Met Val Ser Val Val Leu Ile Phe Ser Phe Leu Ser Leu Thr Ile Ser
1 5 10 15

Thr Thr Ala Ser Ala Tyr Asn Gly Asn Asp Thr Gln Gly Trp Asn Asp
20 25 30

20 Lys Phe His Xaa Xaa Ser Val Lys Thr Gln Thr Xaa
35 40

25

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

35 Met Ile Ser Asp Ala Gly Ala Gly Phe Gly Val Phe Leu Leu Val Pro
1 5 10 15

Arg Ala Gly His Cys Trp Gly Ala Gly Lys Pro Leu Pro Ser Cys Pro
20 25 30

40 Ser Val Ala Ser Ile Pro Ser Trp Val Leu Pro Ser Phe Leu Glu Arg
35 40 45

Gly Arg Xaa
50

45

(2) INFORMATION FOR SEQ ID NO: 244:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

55 Met Val Gln Thr Ile Gln Asp Phe Leu Ser Leu Phe Ser Thr Pro Ile
1 5 10 15

60 Phe Leu Leu Leu Leu Met Phe Glu Thr Leu Ser Leu Ala Pro Ala Trp
20 25 30

Leu Lys Pro Leu Arg Val Thr Ser His Ser Xaa
35 40

5

(2) INFORMATION FOR SEQ ID NO: 245:

- 10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

15 Met Ile Leu Met Pro Gly Leu Gly Thr Ser Arg Gln Arg Arg Ser Val Pro
1 5 10 15

Phe Val Pro Thr Leu Asn Ala Ser Thr Pro Gly Ala Met Thr Gly Pro
20 25 30

20 Thr Ala Thr Leu Thr Ser Cys Gln Trp Thr Thr Ala Cys Arg Val Ser
35 40 45

25 Trp Ala Asn Gly Trp Thr Ser Leu Arg Thr Phe Arg Xaa
50 55 60

30 (2) INFORMATION FOR SEQ ID NO: 246:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Ser His His Ala Gln Pro Arg Phe Leu Leu Ile Thr Met Leu Leu
1 5 10 15

40 Gln Glu Ala Lys Pro Val Ser Asn Ile Pro His Leu Leu Glu Ser Trp
20 25 30

Tyr Phe Gly Xaa
35

45

(2) INFORMATION FOR SEQ ID NO: 247:

- 50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55 Met Asn Ser Leu Phe Trp Met Ile Leu Leu Pro Val Ser Gln Asp Gln
1 5 10 15

60 Val Val Glu Gly Leu Gln Gly Phe Ser Gln Ile His Met Arg Ile
20 25 30

Leu Arg Lys His Leu Xaa
35

5

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 211 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

15 Met Ser Arg Ser Xaa Asp Val Thr Asn Thr Thr Phe Leu Leu Met Ala
1 5 10 15

Ala Ser Ile Tyr Leu His Asp Gln Asn Pro Asp Ala Ala Leu Arg Ala
20 25 30

20 Leu His Gln Gly Asp Ser Leu Glu Cys Thr Ala Met Thr Val Gln Ile
35 40 45

25 Leu Leu Lys Leu Asp Arg Leu Asp Leu Ala Arg Lys Glu Leu Lys Arg
50 55 60

Met Gln Asp Leu Asp Glu Asp Ala Thr Leu Thr Gln Leu Ala Thr Ala
65 70 75 80

30 Trp Val Ser Leu Ala Thr Gly Gly Glu Lys Leu Gln Asp Ala Tyr Tyr
85 90 95

Ile Phe Gln Glu Met Ala Asp Lys Cys Ser Pro Thr Leu Leu Leu
100 105 110

35 Asn Gly Gln Ala Ala Cys His Met Ala Gln Gly Arg Trp Glu Ala Ala
115 120 125

40 Glu Gly Leu Leu Gln Glu Ala Leu Asp Lys Asp Ser Gly Tyr Pro Glu
130 135 140

Thr Leu Val Asn Leu Ile Val Leu Ser Gln His Leu Gly Lys Pro Pro
145 150 155 160

45 Glu Val Thr Asn Arg Tyr Leu Ser Gln Leu Lys Asp Ala His Arg Ser
165 170 175

His Pro Phe Ile Lys Glu Tyr Gln Ala Lys Glu Asn Asp Phe Asp Arg
180 185 190

50 Leu Val Leu Gln Tyr Ala Pro Ser Ala Glu Ala Gly Pro Glu Leu Ser
195 200 205

55 Gly Pro Xaa
210

(2) INFORMATION FOR SEQ ID NO: 249:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 548 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

Met Glu Asp Ser Glu Ala Leu Gly Phe Glu His Met Gly Leu Asp Pro
1 5 10 15

10 Arg Leu Leu Gln Ala Val Thr Asp Leu Gly Trp Ser Arg Pro Thr Leu
20 25 30

Ile Gln Glu Lys Ala Ile Pro Leu Ala Leu Glu Gly Lys Asp Leu Leu
35 40 45

15 Ala Arg Ala Arg Thr Gly Ser Gly Lys Thr Ala Ala Tyr Ala Ile Pro
50 55 60

20 Met Leu Gln Leu Leu His Arg Lys Ala Thr Gly Pro Val Val Glu
65 70 75 80

Gln Ala Val Arg Gly Leu Val Leu Val Pro Thr Lys Glu Leu Ala Arg
85 90 95

25 Gln Ala Gln Ser Met Ile Gln Gln Leu Ala Thr Tyr Cys Ala Arg Asp
100 105 110

Val Arg Val Ala Asn Val Ser Ala Ala Glu Asp Ser Val Ser Gln Arg
115 120 125

30 Ala Val Leu Met Glu Lys Pro Asp Val Val Val Gly Thr Pro Ser Arg
130 135 140

Ile Leu Ser His Leu Gln Gln Asp Ser Leu Lys Leu Arg Asp Ser Leu
35 145 150 155 160

Glu Leu Leu Val Val Asp Glu Ala Asp Leu Leu Phe Ser Phe Gly Phe
165 170 175

40 Glu Glu Glu Leu Lys Ser Leu Leu Cys His Leu Pro Arg Ile Tyr Gln
180 185 190

Ala Phe Leu Met Ser Ala Thr Phe Asn Glu Asp Val Gln Ala Leu Lys
195 200 205

45 Glu Leu Ile Leu His Asn Pro Val Thr Leu Lys Leu Gln Glu Ser Gln
210 215 220

50 Leu Pro Gly Pro Asp Gln Leu Gln Gln Phe Gln Val Val Cys Glu Thr
225 230 235 240

Glu Glu Asp Lys Phe Leu Leu Leu Tyr Ala Leu Leu Lys Leu Ser Leu
245 250 255

55 Ile Arg Gly Lys Ser Leu Leu Phe Val Asn Thr Leu Glu Arg Ser Tyr
260 265 270

Arg Leu Arg Leu Phe Leu Glu Gln Phe Ser Ile Pro Thr Cys Val Leu
275 280 285

60

Asn Gly Glu Leu Pro Leu Arg Ser Arg Cys His Ile Ile Ser Gln Phe
 290 295 300

Asn Gln Gly Phe Tyr Asp Cys Val Ile Ala Thr Asp Ala Glu Val Leu
 5 305 310 315 320

Gly Ala Pro Val Lys Gly Lys Arg Arg Gly Arg Gly Pro Lys Gly Asp
 325 330 335

10 Lys Ala Ser Asp Pro Glu Ala Gly Val Ala Arg Gly Ile Asp Phe His
 340 345 350

His Val Ser Ala Val Leu Asn Phe Asp Leu Pro Pro Thr Pro Glu Ala
 355 360 365

15 Tyr Ile His Arg Ala Gly Arg Thr Ala Arg Ala Asn Asn Pro Gly Ile
 370 375 380

20 Val Leu Thr Phe Val Leu Pro Thr Glu Gln Phe His Leu Gly Lys Ile
 385 390 395 400

Glu Glu Leu Leu Ser Gly Glu Asn Arg Gly Pro Ile Leu Leu Pro Tyr
 405 410 415

25 Gln Phe Arg Met Glu Glu Ile Glu Gly Phe Arg Tyr Arg Cys Arg Asp
 420 425 430

Ala Met Arg Ser Val Thr Lys Gln Ala Ile Arg Glu Ala Arg Leu Lys
 435 440 445

30 Glu Ile Lys Glu Glu Leu Leu His Ser Glu Lys Leu Lys Thr Tyr Phe
 450 455 460

35 Glu Asp Asn Pro Arg Asp Leu Gln Leu Leu Arg His Asp Leu Pro Leu
 465 470 475 480

His Pro Ala Val Val Lys Pro His Leu Gly His Val Pro Asp Tyr Leu
 485 490 495

40 Val Pro Pro Ala Leu Arg Gly Leu Val Arg Pro His Lys Lys Arg Lys
 500 505 510

Lys Leu Ser Ser Ser Cys Arg Lys Ala Lys Arg Ala Lys Ser Gln Asn
 515 520 525

45 Pro Leu Arg Ser Phe Lys His Lys Gly Lys Lys Phe Arg Pro Thr Ala
 530 535 540

50 Lys Pro Ser Xaa
 545

(2) INFORMATION FOR SEQ ID NO: 250:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

Met Thr Thr Val Pro Pro Ser Pro Arg Pro Met Ser Arg Pro Ser Glu
 1 5 10 15

5 Arg Asn Met Arg Arg Pro Arg Gly Pro Ser Pro Leu Pro Ala Ser Pro
 20 25 30

Arg Asn Ser Thr Pro Asp Glu Pro Asp Val His Phe Ser Lys Lys Phe
 35 40 45

10 Leu Asn Val Phe Met Ser Gly Arg Ser Arg Ser Ser Ala Glu Ser
 50 55 60

Phe Gly Leu Phe Ser Cys Ile Ile Asn Gly Glu Glu Gln Glu Gln Thr
 15 65 70 75 80

His Arg Ala Ile Phe Arg Phe Val Pro Arg His Glu Asp Glu Leu Glu
 85 90 95

20 Leu Glu Val Asp Asp Pro Leu Leu Val Glu Leu Gln Ala Glu Asp Tyr
 100 105 110

Trp Tyr Glu Ala Tyr Asn Met Arg Thr Gly Ala Arg Gly Val Phe Pro
 115 120 125

25 Ala Tyr Tyr Ala Ile Glu Val Thr Lys Glu Pro Glu His Met Ala Ala
 130 135 140

Leu Ala Lys Asn Ser Asp Trp Val Asp Gln Phe Arg Val Lys Phe Leu
 30 145 150 155 160

Gly Ser Val Gln Val Pro Tyr His Lys Gly Asn Asp Val Leu Cys Ala
 165 170 175

35 Ala Met Gln Lys Ile Ala Thr Thr Arg Arg Leu Thr Val His Phe Asn
 180 185 190

Pro Pro Ser Ser Cys Val Leu Glu Ile Ser Val Arg Gly Val Lys Ile
 40 195 200 205

Gly Val Lys Ala Asp Asp Ser Gln Glu Ala Lys Gly Asn Lys Cys Ser
 210 215 220

His Phe Phe Gln Leu Lys Asn Ile Ser Phe Cys Gly Tyr His Pro Lys
 45 225 230 235 240

Asn Asn Lys Tyr Phe Gly Phe Ile Thr Lys His Pro Ala Asp His Arg
 245 250 255

50 Phe Ala Cys His Val Phe Val Ser Glu Asp Ser Thr Lys Ala Leu Ala
 260 265 270

Glu Ser Val Gly Arg Ala Phe Gln Gln Phe Tyr Lys Gln Phe Val Glu
 55 275 280 285

Tyr Thr Cys Pro Thr Glu Asp Ile Tyr Leu Glu
 290 295

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

10 Leu Leu Tyr Leu Leu Lys Val Xaa Val Ile Phe Val Phe Ser Ser Ser
1 5 10 15

Lys Gly Val Thr Leu Val Ser Met Asn Leu Thr Ser Phe Phe Val Ser
20 25 30

15 Ser Val Leu Ala Cys Phe Ser Xaa
35 40

20 (2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 594 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

Met Pro Ala Ser Ser Leu Glu Ser Arg Ser Phe Leu Leu Ala Lys Lys
1 5 10 15

Ser Gly Glu Asn Val Ala Lys Phe Ile Ile Asn Ser Tyr Pro Lys Tyr
20 25 30

Phe Gln Lys Asp Ile Ala Glu Pro His Ile Pro Cys Leu Met Pro Glu
35 40 45

Tyr Phe Glu Pro Gln Ile Lys Asp Ile Ser Glu Ala Ala Leu Lys Glu
50 55 60

Arg Ile Glu Leu Arg Lys Val Lys Ala Ser Val Asp Met Phe Asp Gln
60 65 70 75 80

Leu Leu Gln Ala Gly Thr Thr Val Ser Leu Glu Thr Thr Asn Ser Leu
85 90 95

Leu Asp Xaa Leu Cys Tyr Tyr Gly Asp Gln Glu Pro Ser Thr Asp Tyr
100 105 110

His Phe Gln Gln Thr Gly Gln Ser Glu Ala Leu Glu Glu Asn Asp
115 120 125

Glu Thr Ser Arg Arg Lys Ala Gly His Gln Phe Gly Val Thr Trp Arg
130 135 140

55 Ala Lys Asn Asn Ala Glu Arg Ile Phe Ser Leu Met Pro Glu Lys Asn
145 150 155 160

Glu His Ser Tyr Cys Thr Met Ile Arg Gly Met Val Lys His Arg Ala
165 170 175

60

Tyr Glu Gln Ala Leu Asn Leu Tyr Thr Glu Leu Leu Asn Asn Arg Leu
 180 185 190

His Ala Asp Val Tyr Thr Phe Asn Ala Leu Ile Glu Ala Thr Val Cys
 5 195 200 205

Ala Ile Asn Glu Lys Phe Glu Glu Lys Trp Ser Lys Ile Leu Glu Leu
 210 215 220

10 Leu Arg His Met Val Ala Gln Lys Val Lys Pro Asn Leu Gln Thr Phe
 225 230 235 240

Asn Thr Ile Leu Lys Cys Leu Arg Arg Phe His Val Phe Ala Arg Ser
 245 250 255

15 Pro Ala Leu Gln Val Leu Arg Glu Met Lys Ala Ile Gly Ile Glu Pro
 260 265 270

Ser Leu Ala Thr Tyr His His Ile Ile Arg Leu Phe Asp Gln Pro Gly
 20 275 280 285

Asp Pro Leu Lys Arg Ser Ser Phe Ile Ile Tyr Asp Ile Met Asn Glu
 290 295 300

25 Leu Met Gly Lys Arg Phe Ser Pro Lys Asp Pro Asp Asp Asp Lys Phe
 305 310 315 320

Phe Gln Ser Ala Met Ser Ile Cys Ser Ser Leu Arg Asp Leu Glu Leu
 325 330 335

30 Ala Tyr Gln Val His Gly Leu Leu Lys Thr Gly Asp Asn Trp Lys Phe
 340 345 350

35 Ile Gly Pro Asp Gln His Arg Asn Phe Tyr Tyr Ser Lys Phe Phe Asp
 355 360 365

Leu Ile Cys Leu Met Glu Gln Ile Asp Val Thr Leu Lys Trp Tyr Glu
 370 375 380

40 Asp Leu Ile Pro Ser Ala Tyr Phe Pro His Ser Gln Thr Met Ile His
 385 390 395 400

Leu Leu Gln Ala Leu Asp Val Ala Asn Arg Leu Glu Val Ile Pro Lys
 405 410 415

45 Ile Trp Lys Asp Ser Lys Glu Tyr Gly His Thr Phe Arg Ser Asp Leu
 420 425 430

Arg Glu Glu Ile Leu Met Leu Met Ala Arg Asp Lys His Pro Pro Glu
 50 435 440 445

Leu Gln Val Ala Phe Ala Asp Cys Ala Ala Asp Ile Lys Ser Ala Tyr
 450 455 460

55 Glu Ser Gln Pro Ile Arg Gln Thr Ala Gln Asp Trp Pro Ala Thr Ser
 465 470 475 480

Leu Asn Cys Ile Ala Ile Leu Phe Leu Arg Ala Gly Arg Thr Gln Glu
 60 485 490 495

Ala Trp Lys Met Leu Gly Leu Phe Arg Lys His Asn Lys Ile Pro Arg
500 505 510

5 Ser Glu Leu Leu Asn Glu Leu Met Asp Ser Ala Lys Val Ser Asn Ser
515 520 525

Pro Ser Gln Ala Ile Glu Val Val Glu Leu Ala Ser Ala Phe Ser Leu
530 535 540

10 Pro Ile Cys Glu Gly Leu Thr Gln Arg Val Met Ser Asp Phe Ala Ile
545 550 555 560

Asn Gln Glu Gln Lys Glu Ala Leu Ser Asn Leu Thr Ala Leu Thr Ser
565 570 575

15 Asp Ser Asp Thr Asp Ser Ser Asp Ser Asp Ser Asp Thr Ser Glu
580 585 590

20 Gly Lys

25 (2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 131 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

Met Lys Leu Asn Leu Cys Ile Pro Asn Trp Ala Arg Cys Pro Leu Leu
1 5 10 15

35 Leu Leu Phe Pro Gln Leu Leu Pro Phe Gln Gly Glu Asp Asp Asp Pro
20 25 30

Leu Lys Ala Lys Ala Ala Asn Leu Val Glu Ala Val Pro Trp Gly Ile
35 40 45

40 Lys Ala Pro Ser Phe Gln Val Thr Cys Leu Val Arg Val Gln Leu Gln
50 55 60

45 Ser Cys Thr Pro Ser Arg Pro Ser Thr Leu Leu Ala Thr Ser Gln Ser
65 70 75 80

Pro Gly Arg Ile Ser Cys Tyr Ser Pro Leu Ser His Leu Pro Pro Val
85 90 95

50 Thr Thr Ser Ile Gln Pro Ser Pro Val Met Val Pro Phe Gln Tyr Gln
100 105 110

Ala Phe Leu Leu Gln Val Lys Glu Pro Ala Ala Gln Thr Leu Leu Gly
115 120 125

55 Gln Gln Xaa
130

500

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

5 Met Arg Tyr His Ala Gln Leu Ile Phe Cys Ile Phe Cys Xaa Phe Val
10 1 5 10 15
Phe Val Xaa Lys Xaa
20

15

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

25 Met Asn Asp Asn Ser Pro Asn His Ser Ser Ser Tyr Leu Pro Leu Pro
1 5 10 15
Leu Thr Ile Val Ile Leu Gln Thr Gly His Lys Gly Thr Leu Xaa
20 25 30

30

(2) INFORMATION FOR SEQ ID NO: 256:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 219 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

40 Met His Phe Leu Phe Arg Phe Ile Val Phe Phe Tyr Leu Trp Gly Leu
1 5 10 15
Phe Thr Ala Gln Arg Gln Lys Lys Glu Glu Ser Thr Glu Glu Val Lys
45 20 25 30
Ile Glu Val Leu His Arg Pro Glu Asn Cys Ser Lys Thr Ser Lys Lys
35 40 45
50 Gly Asp Leu Leu Asn Ala His Tyr Asp Gly Tyr Leu Ala Lys Asp Gly
50 55 60
Ser Lys Phe Tyr Cys Ser Arg Thr Gln Asn Glu Gly His Pro Lys Trp
55 65 70 75 80
Phe Val Leu Gly Val Gly Gln Val Ile Lys Gly Leu Asp Ile Ala Met
60 85 90 95
Thr Asp Met Cys Pro Gly Glu Lys Arg Lys Val Val Ile Pro Pro Ser
60 100 105 110

501

Phe Ala Tyr Gly Lys Glu Gly Tyr Ala Glu Gly Lys Ile Pro Pro Asp
115 120 125

5 Ala Thr Leu Ile Phe Glu Ile Glu Leu Tyr Ala Val Thr Lys Gly Pro
130 135 140

Arg Ser Ile Glu Thr Phe Lys Gln Ile Asp Met Asp Asn Asp Arg Gln
145 150 155 160

10 Leu Ser Lys Ala Glu Ile Asn Leu Tyr Leu Gln Arg Glu Phe Glu Lys
165 170 175

Asp Glu Lys Pro Arg Asp Lys Ser Tyr Gln Asp Ala Val Leu Glu Asp
15 180 185 190

Ile Phe Lys Lys Asn Asp His Asp Gly Asp Gly Phe Ile Ser Pro Lys
195 200 205

20 Glu Tyr Asn Val Tyr Gln His Asp Glu Leu Xaa
210 215

25 (2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

Met Trp Val Ile Arg Val Phe Gln Lys Thr Phe Leu Phe Phe Val Leu
1 5 10 15

35 Phe Trp Ser Val His Cys Ile Ser Asp Lys Phe Gly Cys Leu Trp His
20 25 30

40 Val Cys Met Lys Arg Glu Gly Asp Xaa Asn Cys Leu Ser Phe Ser Xaa
35 40 45

Leu Xaa
50

45

(2) INFORMATION FOR SEQ ID NO: 258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

55 Met Pro Ser Gln Thr Glu Xaa Phe Ala Ala Cys Gly Gly His Ser Leu
1 5 10 15

Leu Leu Val Xaa Leu Pro Leu Gly Leu Pro Phe Cys Pro Arg Ala Ala
20 25 30

60

Leu Cys Asp Leu Pro Phe Ser Leu Pro Ser Phe Pro Gly Gln Ala Arg
 35 40 45

5 Arg Gly Gly Ala Glu Lys Gln Gly Ala Glu Gly Arg Gly Leu Gln Val
 50 55 60

Lys Pro Arg Gly Gln Arg Thr Phe Gln Val Ser Arg Thr Ala Pro Ala
 65 70 75 80

10 Ala Pro Arg Ser Arg Gln Pro Arg Pro Pro Ala Ala Leu Pro Ala Leu
 85 90 95

Gly Phe Gly Gly Arg Gly Val Ala Lys Gly Arg Phe Leu Cys Phe Trp
 100 105 110

15 Cys Leu Tyr Met Leu Arg Ile Asp Gln Xaa
 115 120

20

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 88 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

Met Thr Ala Phe Cys Ser Leu Leu Leu Gln Ala Gln Ser Leu Leu Pro
 30 1 5 10 15

Arg Thr Met Ala Ala Pro Gln Asp Ser Leu Arg Pro Gly Glu Glu Asp
 20 25 30

35 Glu Gly Met Gln Leu Leu Gln Thr Lys Asp Ser Met Ala Lys Gly Ala
 35 40 45

Arg Pro Gly Ala Xaa Arg Gly Arg Ala Arg Trp Gly Leu Ala Tyr Thr
 50 55 60

40 Leu Leu His Asn Pro Thr Leu Gln Val Phe Arg Lys Thr Ala Leu Leu
 65 70 75 80

45 Gly Ala Asn Gly Ala Gln Pro Xaa
 85

50

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

Met Ile Gln Val Ser Val Pro Leu Leu Thr Ile Met Ile Phe Leu Leu
 1 5 10 15

60 Tyr Leu Gln Ile Gly Pro Gly Lys Leu Xaa

20 25

5 (2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

Met Leu Leu Asp Pro Phe Ile Leu Leu Phe Cys Leu Phe Ser Thr Ala
1 5 10 15

Ala Gln Ser Cys Leu Glu Phe Ile Tyr Ile Gln Phe Xaa
20 25

20

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

Met Lys Phe Leu Ser Ile Leu Leu Asp Asp Asn Asn Phe Xaa Leu Met
30 1 5 10 15

Leu Met Leu Ala Pro Phe Gly Cys Leu Ala Phe Glu Arg Ser Met Lys
20 25 30

35 Met Arg Asn Gly Ala Leu Gly Leu Glu Glu Val Xaa
35 40

40 (2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

Met Arg Thr Leu Phe Asn Leu Leu Trp Leu Ala Leu Ala Cys Ser Pro
1 5 10 15

50 Val His Thr Thr Leu Ser Lys Ser Asp Ala Lys Lys Ala Ala Ser Lys
20 25 30

55 Thr Leu Leu Glu Lys Ser Gln Phe Ser Asp Lys Pro Val Gln Asp Arg
35 40 45

Gly Leu Val Val Thr Asp Leu Lys Ala Glu Ser Val Val Leu Glu His
50 55 60

60 Arg Ser Tyr Cys Ser Ala Lys Ala Arg Asp Arg His Phe Ala Gly Asp

504

65 70 75 80

Val Leu Gly Tyr Val Thr Pro Trp Asn Ser His Gly Tyr Asp Val Thr
85 90 95

5 Lys Val Phe Gly Ser Lys Phe Thr Gln Ile Ser Pro Val Trp Leu Gln
100 105 110

Leu Lys Arg Arg Gly Arg Glu Met Phe Glu Val Thr Gly Leu His Asp
10 115 120 125

Val Asp Gln Gly Trp Met Arg Ala Val Arg Lys His Ala Lys Gly Leu
130 135 140

15 His Ile Val Pro Arg Leu Leu Phe Glu Asp Trp Thr Tyr Asp Asp Phe
145 150 155 160

Arg Asn Val Leu Asp Ser Glu Asp Glu Ile Glu Glu Leu Ser Lys Thr
165 170 175

20 Val Val Gln Val Ala Lys Asn Gln His Phe Asp Gly Phe Val Val Glu
180 185 190

25 Val Trp Asn Gln Leu Leu Ser Gln Lys Arg Val Thr Asp Gln Leu Gly
195 200 205

Met Phe Thr His Lys Glu Phe Glu Gln Leu Ala Pro Val Leu Asp Gly
210 215 220

30 Phe Ser Leu Met Thr Tyr Asp Tyr Ser Thr Ala His Gln Pro Gly Pro
225 230 235 240

Asn Ala Pro Leu Ser Trp Val Arg Ala Cys Val Gln Val Leu Asp Pro
245 250 255

35 Lys Ser Lys Trp Arg Ser Lys Ile Leu Leu Gly Leu Asn Phe Tyr Gly
260 265 270

40 Met Asp Tyr Ala Thr Ser Lys Asp Ala Arg Glu Pro Val Val Gly Ala
275 280 285

Arg Tyr Ile Gln Thr Leu Lys Asp His Arg Pro Arg Met Val Trp Asp
290 295 300

45 Ser Gln Xaa Ser Glu His Phe Phe Glu Tyr Lys Lys Ser Arg Ser Gly
305 310 315 320

Arg His Val Val Phe Tyr Pro Thr Leu Lys Ser Leu Gln Val Arg Leu
325 330 335

50 Glu Leu Ala Arg Glu Leu Gly Val Gly Val Ser Ile Trp Glu Leu Gly
340 345 350

55 Gln Gly Leu Asp Tyr Phe Tyr Asp Leu Leu Xaa
355 360

(2) INFORMATION FOR SEQ ID NO: 264:

60

505

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

Leu Pro Thr Lys Ile Leu Val Lys Pro Asp Arg Thr Phe Glu Ile Lys
1 5 10 15

10 Ile Gly Gln Pro Thr Val Ser Tyr Phe Leu Lys Ala Ala Ala Gly Ile
20 25 30

Glu Lys Gly Ala Arg Gln Thr Gly Lys Glu Val Ala Gly Leu Val Thr
35 40 45

15 Leu Lys His Val Tyr Glu Ile Ala Arg Ile Lys Ala Gln Asp Glu Ala
50 55 60

20 Phe Ala Leu Gln Asp Val Pro Leu Ser Ser Val Val Arg Ser Ile Ile
65 70 75 80

Gly Ser Ala Arg Ser Leu Gly Ile Arg Val Val Lys Asp Leu Ser Ser
85 90 95

25 Glu Glu Leu Ala Ala Phe Gln Lys Glu Arg Ala Ile Phe Leu Ala Ala
100 105 110

Gln Lys Glu Ala Asp Leu Ala Ala Gln Glu Glu Ala Ala Lys Lys Xaa
115 120 125

30

35

(2) INFORMATION FOR SEQ ID NO: 265:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

45 Met Leu Leu Gln Ile His Pro Leu Leu Pro Ser Pro Thr Ile Pro His
1 5 10 15

Ile Leu Leu Leu Phe Leu Tyr Pro Thr Phe Ser Ile Leu Glu His Ser
20 25 30

50 Cys Ser Tyr Cys Ile Glu Tyr Leu Trp Val Cys Leu Leu Phe Cys Leu
35 40 45

Ser Leu Trp Phe Leu Xaa
50

55

(2) INFORMATION FOR SEQ ID NO: 266:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5

Met Cys Leu Trp Cys Cys Gly Asp Val Cys Ser Gly Leu Ser Ser Leu
1 5 10 15

10

Leu Ser Leu Cys Val Cys Cys Val Val Leu Ala Val Cys
20 25

15

(2) INFORMATION FOR SEQ ID NO: 267:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

25

Glu Gly Leu Arg Leu Leu Ser Leu Pro Ala Ala Leu Pro Arg Ser
1 5 10 15

30

Cys Cys His Pro Arg Trp Leu Pro Val Xaa
20 25

35

(2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 221 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

40

Met Phe His Gly Ile Pro Ala Thr Pro Gly Ile Gly Ala Pro Gly Asn
1 5 10 15

Lys Pro Glu Leu Tyr Glu Glu Val Lys Leu Tyr Lys Asn Ala Arg Glu
20 25 30

45

Arg Glu Lys Tyr Asp Asn Met Ala Glu Leu Phe Ala Val Val Lys Thr
35 40 45

Met Gln Ala Leu Glu Lys Ala Tyr Ile Lys Asp Cys Val Ser Pro Ser
50 55 60

50

Glu Tyr Thr Ala Ala Cys Ser Arg Leu Leu Val Gln Tyr Lys Ala Ala
65 70 75 80

Phe Arg Gln Val Gln Gly Ser Glu Ile Ser Ser Ile Asp Glu Phe Cys
85 90 95

55

Arg Lys Phe Arg Leu Asp Cys Pro Leu Ala Met Glu Arg Ile Lys Glu
100 105 110

60

Asp Arg Pro Ile Thr Ile Lys Asp Asp Lys Gly Asn Leu Asn Arg Cys
115 120 125

Ile Ala Asp Val Val Ser Leu Phe Ile Thr Val Met Asp Lys Leu Arg
130 135 140

5 Leu Glu Ile Arg Ala Met Asp Glu Ile Gln Pro Asp Leu Arg Glu Leu
145 150 155 160

Met Glu Thr Met His Arg Met Ser His Leu Pro Pro Asp Phe Glu Gly
165 170 175

10 Arg Gln Thr Val Ser Gln Trp Leu Gln Thr Leu Ser Gly Met Ser Ala
180 185 190

15 Ser Asp Glu Leu Asp Asp Ser Gln Val Arg Gln Met Leu Phe Asp Leu
195 200 205

Glu Ser Ala Tyr Asn Ala Phe Asn Arg Phe Leu His Ala
210 215 220

20

(2) INFORMATION FOR SEQ ID NO: 269:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

30 Met Lys Xaa
1

35 (2) INFORMATION FOR SEQ ID NO: 270:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

45 Met Gln Ala Pro Phe Xaa His Phe Ser Phe Arg Met Phe Ser Asn Leu
1 5 10 15
Tyr Cys Phe Ser Asp Phe Gln Pro Asn Ile Ser Pro Cys Pro Leu Cys
20 25 30
His Cys Ile Leu Pro Xaa His His Val Phe Leu Leu Ala Val
50 35 40 45
Xaa

55

(2) INFORMATION FOR SEQ ID NO: 271:

60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

5 Met Lys Leu Val Thr Met Phe Asp Lys Leu Ser Arg Asn Arg Val Ile
1 5 10 15

Gln Pro Met Gly Met Ser Pro Arg Gly His Leu Thr Ser Leu Gln Asp
20 25 30

10 Ala Met Cys Glu Thr Met Glu Gln Gln Leu Ser Ser Asp Pro Asp Ser
35 40 45

Asp Pro Asp Xaa
15 50

(2) INFORMATION FOR SEQ ID NO: 272:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

Met Ala Val Gly Glu Ala Val Phe Val Pro Leu Gln His Pro Pro Leu
1 5 10 15

30 Leu His Gly Ser Pro Ile Pro Lys Leu Leu Pro Gly Pro Leu Leu Xaa
20 25 30

35

(2) INFORMATION FOR SEQ ID NO: 273:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

Met Asn Gly Cys His Arg Arg Lys Arg Leu His Leu Cys Lys Thr Ile
1 5 10 15

Tyr Leu Leu Trp Phe Val Phe Ser Phe Leu Leu Ser Asn Glu Val Val
20 25 30

50 Ser Ser His Trp His Ile Leu Arg Ala Val Gln Ile Ile Cys Thr Leu
35 40 45

55 Phe His Arg Xaa Ile Ser Ala Phe Xaa
50 55

60 (2) INFORMATION FOR SEQ ID NO: 274:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

Met Gly Trp Val Ser Ser Pro His Val Lys Arg Arg Glu Cys Val Leu
1 5 10 15
10 Lys Lys Pro Phe Phe Xaa
20

15 (2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

25 Met Phe Asn Phe Phe Lys Asn Pro Leu Leu Thr Cys Leu Phe Ile Ser
1 5 10 15
Cys Tyr Leu Tyr Leu Ser Leu Leu Val Asn Lys Val Leu Phe Ala Glu
20 25 30
30 Glu Gly Leu Cys Cys Thr Tyr Cys Thr Thr Ser Asn Thr Gly Glu Gly
35 40 45
Gly Val Xaa
50

35

(2) INFORMATION FOR SEQ ID NO: 276:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

45 Met Xaa
1

50 (2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

60 Met Leu Cys Thr Ile Leu Thr Val Val Ile Ile Ile Ala Ala Gln Thr
1 5 10 15

510

Thr Arg Thr Thr Gly Ile Pro Lys Asn Ala Pro Gly Pro Ala Pro Leu
 20 25 30

5 Cys Ala Pro Arg Ser Pro Arg Leu Phe Leu Gln Xaa Tyr Arg Gly Pro
 35 40 45

Asn Gly Arg Pro Ala His Pro Phe Leu Gly Pro Ser Asp Leu Asp Thr
 50 55 60

10 Ser Xaa
 65

15 (2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 257 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

25 Met Leu Gly Ala Lys Pro His Trp Leu Pro Gly Pro Leu His Ser Pro
 1 5 10 15

Gly Leu Pro Leu Val Leu Val Leu Ala Leu Gly Ala Gly Trp Ala
 20 25 30

30 Gln Glu Gly Ser Glu Pro Val Leu Leu Glu Gly Glu Cys Leu Val Val
 35 40 45

Cys Glu Pro Gly Arg Ala Ala Ala Gly Gly Pro Gly Gly Ala Ala Leu
 50 55 60

35 Gly Glu Ala Pro Pro Gly Arg Val Ala Phe Xaa Ala Val Arg Ser His
 65 70 75 80

40 His His Glu Pro Ala Gly Glu Thr Gly Asn Gly Thr Ser Gly Ala Ile
 85 90 95

Tyr Phe Asp Gln Val Leu Val Asn Glu Gly Gly Phe Asp Arg Ala
 100 105 110

45 Ser Gly Ser Phe Val Ala Pro Val Arg Gly Val Tyr Ser Phe Arg Phe
 115 120 125

His Val Val Lys Val Tyr Asn Arg Gln Thr Val Gln Val Ser Leu Met
 130 135 140

50 Leu Asn Thr Trp Pro Val Ile Ser Ala Phe Ala Asn Asp Pro Asp Val
 145 150 155 160

55 Thr Arg Glu Ala Ala Thr Ser Ser Val Leu Leu Pro Leu Asp Pro Gly
 165 170 175

Asp Arg Val Ser Leu Arg Leu Arg Arg Gly Xaa Ser Thr Gly Trp Leu
 180 185 190

60 Glu Ile Leu Lys Phe Leu Trp Leu Pro His Leu Pro Ser Leu Lys Asp

BEST AVAILABLE COPY

195	200	205
Pro Ser Leu Ser Ser Thr Arg Ile Gln Pro Leu Thr Thr Phe Phe Cys		
210	215	220
5 Pro Leu Leu Pro Xaa Lys Gln Xaa Lys Gln Xaa Xaa Ser Leu Trp		
225	230	235
Leu Leu Ser His Leu Phe Ala Trp Glu Pro Val Pro Asn Thr Gln Val		
10	245	250
Xaa		

15

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:		
20	(A) LENGTH: 103 amino acids	
	(B) TYPE: amino acid	
	(C) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:		
25	Met Ala Pro Arg Ala Leu Pro Gly Ser Ala Val Leu Ala Ala Ala Val	
	1	5
	10	15
Phe Val Gly Gly Ala Val Ser Ser Pro Leu Val Ala Pro Asp Asn Gly		
30	20	25
	30	
Ser Ser Arg Thr Leu His Ser Arg Thr Glu Thr Thr Pro Ser Pro Ser		
	35	40
	45	
Asn Asp Thr Gly Asn Gly His Pro Glu Tyr Ile Ala Tyr Ala Leu Val		
35	50	55
	60	
Pro Val Phe Phe Ile Met Gly Leu Phe Gly Val Leu Ile Xaa Pro Xaa		
	65	70
	75	80
40	Xaa Xaa Lys Lys Gly Tyr Arg Cys Thr Thr Glu Ala Glu Gln Asp	
	85	90
	95	
Ile Glu Glu Glu Lys Gly Xaa		
45	100	

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:		
50	(A) LENGTH: 33 amino acids	
	(B) TYPE: amino acid	
	(C) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:		
55	Met Pro Val Thr Leu Ser Ser Leu Gly Phe Trp Val Leu Leu Ser Leu	
	1	5
	10	15
Leu Phe Pro Trp Arg Thr Asp Gln Gly Cys Gly Pro Ala Thr Cys Tyr		
60	20	25
	30	

xaa

5

(2) INFORMATION FOR SEQ ID NO: 281:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEO ID NO: 281:

15 Met Val Leu Gly Leu Leu Leu Leu Xaa Phe Phe Ser Phe Ser Ser
 1 5 10 15

 Ser Pro Ser Pro Ser Ser Ser Leu Leu Leu Leu Ser Ser Phe Phe Phe
 20 25 30

 20 Gln Ser Leu Ala Leu Ser Pro Arg Leu Glu Xaa
 35 40

25 (2) INFORMATION FOR SEQ ID NO: 282:

35 Glu Trp Leu Val Phe Thr Phe Leu Leu Val Phe Gly Ser Pro Leu Gly
 1 5 10 15

 Lys Gly Pro Leu Xaa
 20

(2) INFORMATION FOR SEQ ID NO: 283:

50	Met Ile Arg Ala Leu Ser Leu Phe Leu Leu Ile Phe Asp Ala Ala Leu
	1 5 10 15
55	Phe Ser Leu Ser Val Phe Val Phe Ile Gly His Leu Leu Pro Met Pro
	20 25 30
60	Lys Gly Thr Gly Leu His Ser Cys Ala Lys His Leu Ile Lys Ser Leu
	35 40 45
65	Lys Glu Asn Val Leu Pro Leu Met Asn Tyr Pro Asp Cys Lys Leu Lys
	50 55 60

Ile Asn Ile Ser Pro Xaa
65 70

5

(2) INFORMATION FOR SEQ ID NO: 284:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

15 Met Gly Lys Leu Ile Arg Leu Ser Val Met Val Met Ser Val Arg Arg
1 5 10 15

Leu Phe Ser Ile Tyr Trp Val Leu Ser Thr Val Pro Asp Ala Val Gly
20 25 30

20 Ser Arg Gly Gly Met Glu Glu Cys Ser Arg Gly Leu Cys Cys Val
35 40 45

25 Ala Gly Gln His Lys Gln Ala Lys Gly Lys Arg Gln Ala Trp Asn Lys
50 55 60

Gly Gly Glu Tyr Gln Cys Val Thr Tyr Cys Xaa
65 70 75

30

(2) INFORMATION FOR SEQ ID NO: 285:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

40 Met Pro Ala Leu Val Thr Leu Leu Leu Phe Pro Leu Leu Pro Leu
1 5 10 15

Met Glu Ala Ser Cys His Val Met Arg Cys Pro Met Glu Arg Pro Thr
20 25 30

45 Xaa

50

(2) INFORMATION FOR SEQ ID NO: 286:

55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

60 Glu Ala Pro Trp Gly Leu Leu Lys Leu Leu Leu Leu Ala Val Phe
1 5 10 15

Xaa

5

(2) INFORMATION FOR SEQ ID NO: 287:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

15 Met Gln Gln Lys Gln Lys Lys Ala Asn Glu Lys Lys Glu Glu Pro Lys
1 5 10 15

Xaa

20

(2) INFORMATION FOR SEQ ID NO: 288:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

30 Met Gln Arg Lys Val Ser Asp Phe Ile Ile His Gln Arg Leu Thr Val
1 5 10 1535 Asn Leu Cys Val Ile Ser Phe Phe Phe Leu Pro Ile Cys Ile Phe
20 25 30Ser Leu Ala Lys Lys Xaa
35

40

(2) INFORMATION FOR SEQ ID NO: 289:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

50 Met Ala Leu Leu Ile Ser Ser Leu Ile Trp Ser Xaa
1 5 10

55 (2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

Met Gln Met Phe Thr Val Ser Leu Leu Leu Ser Leu Leu Leu Arg Ser
1 5 10 15
5 Thr Asp Gln Asn His Leu Gln Leu Leu Val Gly Arg Glu Asp His Tyr
20 25 30
10 Gly Gly Xaa
35

(2) INFORMATION FOR SEQ ID NO: 291:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

Met Ser Glu Ser Ala Cys Ile Leu Asn Asn Gln Lys Glu Leu Xaa
1 5 10 15

25 (2) INFORMATION FOR SEQ ID NO: 292:
(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

35 Met Asp Leu Asp Arg Val Lys Ala Glu Ala Thr Glu Asp Ile Thr Ser
1 5 10 15
Gly Val Leu Cys Leu Leu Phe Leu Arg Leu Pro Pro Asn Ser Cys Ile
20 25 30
40 Phe Pro Ser Ala Val Leu Gly Ser Thr Arg Thr Xaa
35 40

45 (2) INFORMATION FOR SEQ ID NO: 293:
(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 136 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

55 Val Val Gly Thr Gly Thr Ser Leu Ala Leu Ser Ser Leu Leu Ser Leu
1 5 10 15
Leu Leu Phe Ala Gly Met Gln Met Tyr Ser Arg Gln Leu Ala Ser Thr
20 25 30
60 Glu Trp Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe Val

516

	35	40	45
5	Phe Ser Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe Gly Lys 50	55	60
	Gly Phe Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu Leu 65	70	75
10	Ala Leu Phe Ala Ser Gly Leu Ile His Arg Val Cys Val Thr Thr Cys 85	90	95
	Phe Ile Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser Ser 100	105	110
15	Thr Leu Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala Lys Val Thr 115	120	125
20	Gly Lys Ser Lys Lys Arg Asn Xaa 130	135	

(2) INFORMATION FOR SEQ ID NO: 294:

Ile His Asn Tyr Cys Val Leu Asp Lys Leu Arg Asp Phe Val Ala Ser
65 70 75 80

5 Pro Pro Cys Trp Lys Val Ala Gln Val Asp Ser Leu Lys Asp Lys Ala
85 90 95

Arg Lys Leu Tyr Thr Ile Met Asn Ser Phe Cys Arg Arg Asp Leu Val
100 105 110

10 Phe Leu Leu Asp Asp Cys Asn Ala Leu Glu Tyr Pro Ile Pro Val Thr
115 120 125

15 Thr Val Leu Pro Asp Arg Gln Arg Xaa
130 135

20 (2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 58 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

Met Trp Leu Leu Lys Pro Ser Ala His Ser Pro Val His Xaa Leu Val
1 5 10 15

30 Leu Leu Phe Pro Arg Gly Trp Ser Gln Pro Gly Thr His Lys Arg Gln
20 25 30

35 Ile Leu Val Asn Xaa Ala Ser Leu Pro Gly Gly Cys Leu Leu Pro Trp
35 40 45

35 Ile Trp Ser Gly Ala Ala Leu Arg Phe Xaa
50 55

40 (2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

50 Met Ser Arg Arg Ala Glu Ala Ser Ile Phe Val Leu Pro Lys Thr Leu
1 5 10 15

55 Leu Phe Val Leu Phe Pro Ala Phe Pro Ser Pro Ala Val Gly Cys Pro
20 25 30

55 Val Pro Xaa
35

60 (2) INFORMATION FOR SEQ ID NO: 298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

Ser Cys Tyr Ile Thr Pro Trp Ser Lys Ile Gln Ser Phe Ser Leu Ser
1 5 10 15

10 Leu Phe Gln Phe Ile Leu Gln Glu Val Asn Ile Thr Leu Pro Glu Asn
20 25 30

15 Ser Val Trp Tyr Glu Arg Tyr Lys Phe Asp Ile Pro Val Phe His Leu
35 40 45

Asn Gly Gln Phe Leu Met Met His Arg Val Asn Thr Ser Lys Leu Glu
50 55 60

20 Lys Gln Leu Leu Lys Leu Glu Gln Gln Ser Thr Gly Xaa Xaa
65 70 75

25 (2) INFORMATION FOR SEQ ID NO: 299:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

Met Phe Val Leu Phe Ser Leu Pro Lys Tyr Ala Gly Leu Arg Leu Pro
1 5 10 15

35 Ile Pro Gly Leu Ser Ala Leu Leu Val Phe Leu Leu Ser Leu Phe Ser
20 25 30

40 Arg Arg Ala Gln Val Glu Leu Thr Thr Gly Arg Glu Thr Leu Pro Lys
35 40 45

Asn Leu Gln Gly Tyr Phe Pro Glu Phe Gly Phe Gln Val Gln Asn Phe
50 55 60

45 Leu Ser Cys Lys Ile Tyr Ala Ala Ser Gln Lys Gln Pro Leu Pro Pro
65 70 75 80

50 Leu Tyr Gln Leu Arg Phe Tyr Leu Lys His Met Gly Leu Pro Xaa
85 90 95

(2) INFORMATION FOR SEQ ID NO: 300:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

Met Ser Ser His Trp Thr Leu Lys Ile Leu Leu Val Pro Leu Phe Tyr
1 5 10 15

5 Leu Ser Leu Glu Phe Pro Ser Gly Phe Val Leu Cys Leu Ala Asn Asp
20 25 30

Leu Gly Tyr His Phe Ser Ser Arg Val Arg Ser Xaa
35 40

10

(2) INFORMATION FOR SEQ ID NO: 301:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

20 Met Leu Val Val Asn Ile Asn Leu Val Phe Leu Leu Phe Phe Ile Phe
1 5 10 15

25 Leu Cys Tyr Leu Asp Ala Cys Ile Asn Val Phe Cys Phe Tyr Xaa
20 25 30

25

(2) INFORMATION FOR SEQ ID NO: 302:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

35 Met Pro Val Leu Pro Gly Arg Thr Thr Ala Leu Leu Ser Leu Thr Leu
1 5 10 15

40 Ala Phe Ala Val Pro Cys Ser Gly Val Glu Ala Gly Pro Cys Val Pro
20 25 30

Arg Ser His Gly Cys Ser Ser Trp Glu Ala Ser Val Cys Val Thr Ser
35 40 45

45 Ser Thr Pro Gly Gly Ser Trp Arg Ala Arg Ala Leu Phe Pro Ser Ala
50 55 60

50 Ala Trp His Arg Xaa Ala Ala Trp Asp Ser Pro Trp Thr Gln Thr Gly
65 70 75 80

Asp Phe Ala Arg Gly Ala Met Gly Gly Ala Gly Ala Leu Pro Gly Gly
85 90 95

55 Cys Val Cys Ile Ser Gly Arg Pro Arg Ala Gln Lys Leu Pro Ala Leu
100 105 110

Xaa

60

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

10 Thr His Ile His Thr His Ile Ile Ile Cys Ser Ser Val Xaa
1 5 10

15 (2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

Met Glu Asn Phe Phe Phe Ser Phe Tyr Leu Phe Leu Ile Thr Leu Ile
1 5 10 15

20 Pro Asn Gly Arg Thr Leu Ser Thr Thr Ala Asp His Cys Lys Ile Pro
20 25 30

30 Cys Ile Xaa
35

35 (2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

Met Glu Leu Trp Glu Leu Ala Leu Cys Leu Leu Val Ala Leu Ser Ala
1 5 10 15

45 His Met Phe Thr Val Gln Leu Leu Ala Asp Leu Gly Phe Leu Phe Gly
20 25 30

Gly Phe Xaa
35

50

55 (2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 82 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

60

521

Met Gly Ala Gly Ile Leu Ala Leu Leu Pro Leu Glu Ser Val Leu
 1 5 10 15

5 Thr Cys Ser Trp Ile Ser Val Ser Thr Ser Glu Arg Gln Leu Trp Gln
 20 25 30

Ser Ser Gln Lys Ala Thr Ile Leu Ser Leu Lys Leu Asp Ser Cys Phe
 35 40 45

10 Cys Gly His Ser Gly Leu Lys Gly Lys Asn Glu Asp Thr Asp Ser Ser
 50 55 60

Val Pro Ile Ile Pro Ser Lys Thr His Thr His Leu Gly Lys His Leu
 65 70 75 80

15 Ile Xaa

20

(2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 72 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

30 Met Phe Tyr Phe Val Leu Phe Ile Tyr Ser Ser Ser Glu Thr Trp Ser
 1 5 10 15

Gly Ser Val Ala Gln Asp Gly Val His Gly Val Ile Ile Gly His Cys
 20 25 30

35 Ser Val Glu Leu Pro Gly Ser Gly Asp Pro Pro Ala Ser Ala Xaa Leu
 35 40 45

Val Ala Gly Thr Ile Gly Tyr Cys Pro Thr Met Pro Gly Phe Val Tyr
 50 55 60

40 Phe Leu Asn Asp Val Xaa Asn Zaa
 65 70

45

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

55 Met Asp Ser Thr Leu Arg Gln Gly Arg Xaa Leu Leu Thr Leu Val Pro
 1 5 10 15

Ala Ser Leu Phe Ser Leu Thr Leu Gly Gly Pro Gly Pro Trp Lys Asp
 20 25 30

60 Pro Xaa

5 (2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Met Gln Val Val Gly Ser Trp Pro Gly Arg Val Gly Val Val Gly Leu
1 5 10 15

15 Ala Phe Ser Leu Val Ile Pro Pro Pro Ala Ile Cys Ile Ala Gly Pro
20 25 30

20 Ala Pro Gly Leu Gly Gly Glu Arg Gln Gln Lys Gly Leu Gly Arg
35 40 45

Gly Gly Gly Leu Arg Asn Cys Pro Gly Arg Val Gly Met Ala Ala
50 55 60

25 Glu Pro Gly Ala Leu Leu Cys Leu Thr Ser Arg Asp Gly Ser Leu Leu
65 70 75 80

Leu Ser Cys Val Arg Pro His His Val Ile Lys Pro Lys Gly Thr Ala
85 90 95

30 Lys Xaa Xaa
100 105 110

35 Gly Gly Xaa
115

40 (2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

Met Asp Leu Pro Gln Phe Ile Tyr Leu Phe Ile Phe Cys Phe Cys Cys
1 5 10 15

50 Leu Ala Ile Val Asn Asn Ala Ser Ile Asn Ile His Ile Gln Val Ser
20 25 30

Met Trp Leu Tyr Val Phe Ile Ser Leu Gly Tyr Leu His Gly Ser Arg
35 40 45

55 Ile Leu Gly His Asn Ile Ile Leu Cys Leu Thr Ser Gln Arg Ile Ala
50 55 60

60 Lys Arg Phe Phe Ile Val Ala Ala Ser Phe Thr Phe Pro Pro Ala Met
65 70 75 80

Tyr Lys Asp Phe Tyr Phe Ser Ile Ser Leu His Leu Pro Thr Leu Leu
85 90 95

5 Phe Xaa Xaa Xaa Phe Val Phe Ser Leu Leu Pro Pro
100 105

10 (2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 65 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

Met Cys Ser Pro Ser Leu Ser Ser Ser Pro Pro Pro Pro Leu Leu Gln Val
1 5 10 15

20 Phe Phe Phe Phe Phe Ser Pro His Trp Ala Ala Lys Val Val Pro
20 25 30

25 Gln Trp Lys Xaa Arg His Pro Gln Val Ser Ser Gln Leu Leu Leu Cys
35 40 45

Phe Leu Arg Val Asn Cys Gln Phe Leu Phe Leu Gln Glu Ile Leu Phe
50 55 60

30 Xaa
65

35 (2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

Met Cys Leu Ser Arg Trp Lys Ile Phe Tyr Thr Leu Leu Ile Leu Phe
1 5 10 15

45 Xaa Xaa Phe Ser Ile Thr Ser Glu Xaa Glu Thr Phe Tyr Met Ile Ile
20 25 30

50 Ile His His Asn Pro Thr Gln Ile Thr Ala Ser Cys Ser Phe Thr Phe
35 40 45

Leu Xaa
50

55

(2) INFORMATION FOR SEQ ID NO: 313:

(i) SEQUENCE CHARACTERISTICS:

60 (A) LENGTH: 293 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

5 Met Glu Arg Pro Asp Trp Glu Thr Ala Ile Gln Lys Pro Leu Cys Ser
1 5 10 15

Leu Pro Ala Gly Ser Gly Asn Ala Leu Ala Ala Ser Leu Asn His Tyr
20 25 30

10 Ala Gly Tyr Xaa Gln Val Thr Asn Glu Asp Leu Leu Thr Asn Cys Thr
35 40 45

15 Leu Leu Leu Cys Arg Arg Leu Leu Ser Pro Met Asn Leu Leu Ser Leu
50 55 60

His Thr Ala Ser Gly Leu Arg Leu Phe Ser Val Leu Ser Leu Ala Trp
65 70 75 80

20 Gly Phe Ile Ala Asp Val Asp Leu Glu Ser Glu Lys Tyr Arg Arg Leu
85 90 95

Gly Glu Met Arg Phe Thr Leu Gly Thr Phe Leu Arg Leu Ala Ala Leu
100 105 110

25 Arg Thr Tyr Arg Gly Arg Leu Ala Tyr Leu Pro Val Gly Arg Val Gly
115 120 125

Ser Lys Thr Pro Ala Ser Pro Val Val Val Gln Gln Gly Pro Val Asp
30 130 135 140

Ala His Leu Val Pro Leu Glu Glu Pro Val Pro Ser His Trp Thr Val
145 150 155 160

35 Val Pro Asp Glu Asp Phe Val Leu Val Leu Ala Leu Leu His Ser His
165 170 175

Leu Gly Ser Glu Met Phe Ala Ala Pro Met Gly Arg Cys Ala Ala Gly
40 180 185 190

Val Met His Leu Phe Tyr Val Arg Ala Gly Val Ser Arg Ala Met Leu
195 200 205

45 Leu Arg Leu Phe Leu Ala Met Glu Lys Gly Arg His Met Glu Tyr Glu
210 215 220

Cys Pro Tyr Leu Val Tyr Val Pro Val Val Ala Phe Arg Leu Glu Pro
225 230 235 240

50 Lys Asp Gly Lys Gly Val Phe Ala Val Asp Gly Glu Leu Met Val Ser
245 250 255

Glu Ala Val Gln Gly Gln Val His Pro Asn Tyr Phe Trp Met Val Ser
55 260 265 270

Gly Cys Val Glu Pro Pro Pro Ser Trp Lys Pro Gln Gln Met Pro Pro
275 280 285

60 Pro Glu Glu Pro Leu
290

(2) INFORMATION FOR SEQ ID NO: 314:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 68 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Pro Leu Glu Gly Phe Cys Leu Val Asp Ile Gly Phe Leu Leu
1 5 10 15

15 Val Met Leu Ile Ser Leu Ala Ser Glu Cys Phe Thr Thr Cys Leu Asp
20 25 30

Ser Phe Ser Thr Thr Glu Pro Gly Cys Lys Phe Tyr Lys Leu Leu His
35 40 45

20 Ser Val Ser Leu Leu Asn Ile Asn Phe Asn Val Lys Ser Leu Leu Cys
50 55 60

25 Ser His Ile Xaa
65

(2) INFORMATION FOR SEQ ID NO: 315:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Met Pro Leu Gln Leu Ser Gly Gln Tyr Trp Ile Ser Leu Leu Val Phe
1 5 10 15

40 Leu Ser Leu Gln Pro Phe Pro Gln Ala Ala Ile Pro Cys Ala Leu Thr
20 25 30

Asp Val Gly Gly Ser Cys Val Ile Cys His Ile Leu Leu Asn Cys Leu
35 40 45

45 Cys Ile Leu Phe Thr Leu Thr Ala Pro Ser Leu Ser His Val Leu Leu
50 55 60

50 Ile Lys Met Ser Leu Ser Val Cys Tyr Glu Pro Gly Ala Asp Leu Ser
65 70 75 80

Asp Arg Ala Ala Thr Gly Asn Lys Leu Thr Arg Ser Thr Cys Leu
85 90 95

55 Leu Met His Ser Asn Lys Leu Cys Xaa
100 105

60 (2) INFORMATION FOR SEQ ID NO: 316:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

Met Trp Gly Cys Ser Gly Leu Gly His Arg Thr Val Ser Phe Leu Leu
1 5 10 15

10 Leu Leu Pro Cys Ser Phe Pro Arg Pro Cys Xaa Leu Phe Gly Leu Ile
20 25 30

15 Pro Ile Ser Arg Pro Cys Lys Val Glu Ala Pro Arg Leu Ser Val Pro
35 40 45

Xaa Leu Ser Cys Ala Ser His Pro Tyr Cys Asn Cys Pro Met Ser Thr
50 55 60

20 Ser Cys Pro Leu Pro Arg Xaa
65 70

25 (2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Met Leu Asn Val Leu Ser Lys Val Gln Gln Leu Val Ser Xaa Leu Gly
1 5 10 15

35 Leu Val Thr Phe Leu Leu Asn His Ser Ala Ala Gly Gly Ser Pro Gln
20 25 30

40 His Arg Trp Leu Leu Leu Xaa
35

45 (2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 72 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

Met Lys Ala Ile Ala Arg Ala Cys Leu Leu Leu Ser Leu Leu Val Leu
1 5 10 15

55 Pro His Val Val Ser Glu His Leu Phe Trp His His Asn Pro Arg His
20 25 30

Pro Val Ile Trp Pro Phe Pro Pro Phe His Leu Ile Ser Cys Ser Val.
35 40 45

Ser Ala Ser Thr Trp His Leu Gly Glu Xaa Leu Leu Leu Val Pro
50 55 60

Ile Ala Pro Ser Val Trp Ser Xaa
5 65 70

10 (2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

Met Glu Gln Gly Gly Gly Pro Arg Leu Leu Leu Ile Pro Gly Leu
1 5 10 15

20 Leu His Asn Thr Tyr Leu Ala Arg Pro Gly Asp Phe Pro Ala Gln Gly
20 25 30

Thr Thr Glu Asn Thr Glu Cys Gln Gly Ser Pro Ser Pro Ile Ser His
35 40 45

25 Leu Gly Lys Val Arg Ser Leu Asp Ser Asn Thr Gln Ile Xaa
50 55 60

30 (2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 286 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

40 Met Pro Leu Leu Phe Phe Ser Val Ser Thr Leu Phe Ser Gly Ser Val
1 5 10 15

Thr Leu Gln Gln Arg Gly Met Phe Leu Pro Trp Thr Gly Thr Gly Glu
20 25 30

45 Gln Val Leu Ala Leu Leu Trp Pro Arg Phe Glu Leu Ile Leu Glu Met
35 40 45

Asn Val Gln Ser Val Arg Ser Thr Asp Pro Gln Arg Leu Gly Gly Leu
50 55 60

50 Asp Thr Arg Pro His Tyr Ile Thr Arg Arg Tyr Ala Glu Phe Ser Ser
65 70 75 80

55 Ala Leu Val Ser Ile Asn Gln Thr Ile Pro Asn Glu Arg Thr Met Gln
85 90 95

Leu Leu Gly Gln Leu Gln Val Glu Val Glu Asn Phe Val Leu Arg Val
100 105 110

60 Ala Ala Glu Phe Ser Ser Arg Lys Glu Gln Leu Val Phe Leu Ile Asn

	115	120	125
	Asn Tyr Asp Met Met Leu Gly Val Leu Met Glu Arg Ala Ala Asp Asp		
	130	135	140
5	Ser Lys Glu Val Glu Ser Phe Gln Gln Leu Leu Asn Ala Arg Thr Gln		
	145	150	155
	Glu Phe Ile Glu Glu Leu Leu Ser Pro Pro Phe Gly Gly Leu Val Ala		
10	165	170	175
	Phe Val Lys Glu Ala Glu Ala Leu Ile Glu Arg Gly Gln Ala Glu Arg		
	180	185	190
15	Leu Arg Gly Glu Glu Ala Arg Val Thr Gln Leu Ile Arg Gly Phe Gly		
	195	200	205
	Ser Ser Trp Lys Ser Ser Val Glu Ser Leu Ser Gln Asp Val Met Arg		
20	210	215	220
	Ser Phe Thr Asn Phe Arg Asn Gly Thr Ser Ile Ile Gln Gly Ala Leu		
	225	230	235
	Thr Gln Leu Ile Gln Leu Tyr His Arg Phe His Arg Val Leu Ser Gln		
25	245	250	255
	Pro Gln Leu Arg Ala Leu Pro Ala Arg Ala Glu Leu Ile Asn Ile His		
	260	265	270
30	His Leu Met Val Glu Leu Lys Lys His Lys Pro Asn Phe Xaa		
	275	280	285

35 (2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

	Met Phe Arg Ala Leu Arg Asp Leu Leu Thr His Tyr Pro Gln Gln Ile		
	1	5	10
45			15
	Leu Leu Gln Val Leu Val Val Met Tyr Gln Val Leu Gln Val Trp Glu		
	20	25	30
	Leu Pro Trp Pro Glu Leu Ile His Leu Gln Gly Ile Val Pro Thr Asp		
50	35	40	45
	Gln Leu His Leu Lys Gln Xaa		
	50	55	

55

(2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids

60

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

5 Asp Phe Val Pro Val Leu Val Phe Val Leu Ile Lys Ala Asn Pro Pro
1 5 10 15

10 Cys Leu Leu Ser Thr Val Gln Tyr Ile Ser Ser Phe Tyr Ala Ser Cys
20 25 30

Leu Ser Gly Glu Glu Ser Tyr Trp Trp Met Gln Phe Thr Ala Ala Val
35 40 45

15 Glu Phe Ile Lys Thr Ile Asp Asp Arg Lys Xaa
50 55

(2) INFORMATION FOR SEQ ID NO: 323:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

Met His Pro Ala Arg Lys Leu Leu Ser Leu Leu Phe Leu Ile Leu Met
1 5 10 15

30 Gly Thr Glu Leu Thr Gln Asp Ser Ala Ala Pro Asp Ser Leu Leu Arg
20 25 30

Ser Ser Lys Gly Ser Thr Arg Gly Ser Leu Ala Ala Ile Val Ile Trp
35 40 45

35 Arg Gly Lys Ser Glu Ser Arg Ile Ala Lys Thr Pro Gly Ile Phe Arg
50 55 60

40 Gly Gly Gly Thr Leu Val Leu Pro Pro Thr His Thr Pro Glu Trp Leu
65 70 75 80

Ile Leu Pro Leu Gly Ile Thr Leu Pro Leu Gly Ala Pro Glu Thr Gly
85 90 95

45 Gly Gly Asp Cys Ala Ala Glu Thr Trp Lys Gly Ser Gln Arg Ala Gly
100 105 110

Gln Leu Cys Ala Leu Leu Ala Xaa
115 120

50

(2) INFORMATION FOR SEQ ID NO: 324:

55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

530

Phe Phe Leu Val Val Phe Ser Leu Ser Phe Xaa Pro Ser Val Leu Thr
1 5 10 15
Ser Pro Val His Xaa Pro His Cys Cys Gln Xaa Asp Xaa Ile Leu Phe
5 20 25 30
Lys Asn Thr Leu Xaa Xaa Phe Xaa Ala Lys Tyr Xaa
35 40

10

(2) INFORMATION FOR SEQ ID NO: 325:

(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 59 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

20 Met Phe Ser Arg Thr Ser Asn Phe Trp Thr Phe Phe Phe Gln Phe Leu
1 5 10 15
Ile Phe Lys Val Phe Leu Val Leu Lys Asn Xaa Phe Thr Ser Gln Lys
25 20 25 30
25 Ile Xaa Xaa Ile Xaa Xaa Glu Lys Pro Lys Lys Lys Xaa Arg Gly
35 40 45
Gly Arg Ala Pro Ser Pro Gln Gly Gly Pro Xaa
30 50 55

(2) INFORMATION FOR SEQ ID NO: 326:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Gly Leu Leu Ile Phe Met Leu Leu Ile Gly Ile His Ser Gln Cys
1 5 10 15

45 Ser Xaa

50 (2) INFORMATION FOR SEQ ID NO: 327:

(i) SEQUENCE CHARACTERISTICS:
55 (A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Val Leu Phe Cys Phe Val Leu Phe Cys Phe Val Phe Glu Met Asp.
1 5 10 15

60

Ser Ser Ser Val Thr Gln Ala Gly Val Gln Trp Cys Asp Leu Gly Ser
 20 25 30

Leu Gln Ala Pro Pro Pro Gly Phe Ser Pro Phe Ser Cys Leu Ser Leu
 5 35 40 45

Pro Ser Ser Trp Asp Tyr Arg Arg Pro Pro Pro Arg Pro Ala Asn Phe
 50 55 60

10 Leu Tyr Phe Leu Val Glu Thr Gly Phe His His Val Ser Gln Asp Gly
 65 70 75 80

Leu Asp Leu Leu Thr Ser Xaa
 15 85

(2) INFORMATION FOR SEQ ID NO: 328:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 538 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

Met Ser Thr Lys Lys Leu Cys Ile Val Gly Gly Ile Leu Leu Val Phe
 1 5 10 15

30 Gln Ile Ile Ala Phe Leu Val Gly Gly Leu Ile Ala Pro Gly Pro Thr
 20 25 30

Thr Ala Val Ser Tyr Met Ser Val Lys Cys Val Asp Ala Arg Lys Asn
 35 40 45

35 His His Lys Thr Lys Trp Phe Val Pro Trp Gly Pro Asn His Cys Asp
 50 55 60

Lys Ile Arg Asp Ile Glu Glu Ala Ile Pro Arg Glu Ile Glu Ala Asn
 65 70 75 80

40 Asp Ile Val Phe Ser Val His Ile Pro Leu Pro His Met Glu Met Ser
 85 90 95

45 Pro Trp Phe Gln Phe Met Leu Phe Ile Leu Gln Leu Asp Ile Ala Phe
 100 105 110

Lys Leu Asn Asn Gln Ile Arg Glu Asn Ala Glu Val Ser Met Asp Val
 115 120 125

50 Ser Leu Ala Tyr Arg Asp Asp Ala Phe Ala Glu Trp Thr Glu Met Ala
 130 135 140

His Glu Arg Val Pro Arg Lys Leu Lys Cys Thr Phe Thr Ser Pro Lys
 145 150 155 160

55 Thr Pro Glu His Glu Gly Arg Tyr Tyr Glu Cys Asp Val Leu Pro Phe
 165 170 175

60 Met Glu Ile Gly Ser Val Ala His Lys Phe Tyr Leu Leu Asn Ile Arg
 180 185 190

Leu Pro Val Asn Glu Lys Lys Ile Asn Val Gly Ile Gly Glu Ile
 195 200 205

5 Lys Asp Ile Arg Leu Val Gly Ile His Gln Asn Gly Gly Phe Thr Lys
 210 215 220

Val Trp Phe Ala Met Lys Thr Phe Leu Thr Pro Ser Ile Phe Ile Ile
 225 230 235 240

10 Met Val Trp Tyr Trp Arg Arg Ile Thr Met Met Ser Arg Pro Pro Val
 245 250 255

Leu Leu Glu Lys Val Ile Phe Ala Leu Gly Ile Ser Met Thr Phe Ile
 15 260 265 270

Asn Ile Pro Val Glu Trp Phe Ser Ile Gly Phe Asp Trp Thr Trp Met
 275 280 285

20 Leu Leu Phe Gly Asp Ile Arg Gln Gly Ile Phe Tyr Ala Met Leu Leu
 290 295 300

Ser Phe Trp Ile Ile Phe Cys Gly Glu His Met Met Asp Gln His Glu
 305 310 315 320

25 Arg Asn His Ile Ala Gly Tyr Trp Lys Gln Val Gly Pro Ile Ala Val
 325 330 335

Gly Ser Phe Cys Leu Phe Ile Phe Asp Met Cys Glu Arg Gly Val Gln
 30 340 345 350

Leu Thr Asn Pro Phe Tyr Ser Ile Trp Thr Thr Asp Ile Gly Thr Glu
 355 360 365

35 Leu Ala Met Ala Phe Ile Ile Val Ala Gly Ile Cys Leu Cys Leu Tyr
 370 375 380

Phe Leu Phe Leu Cys Phe Met Val Phe Gln Val Phe Arg Asn Ile Ser
 385 390 395 400

40 Gly Lys Gln Ser Ser Leu Pro Ala Met Ser Lys Val Arg Arg Leu His
 405 410 415

Tyr Glu Gly Leu Ile Phe Arg Phe Lys Phe Leu Met Leu Ile Thr Leu
 45 420 425 430

Ala Cys Ala Ala Met Thr Val Ile Phe Phe Ile Val Ser Gln Val Thr
 435 440 445

50 Glu Gly His Trp Lys Trp Gly Gly Val Thr Val Gln Val Asn Ser Ala
 450 455 460

Phe Phe Thr Gly Ile Tyr Gly Met Trp Asn Leu Tyr Val Phe Ala Leu
 465 470 475 480

55 Met Phe Leu Tyr Ala Pro Ser His Lys Asn Tyr Gly Glu Asp Gln Ser
 485 490 495

Asn Gly Met Gln Leu Pro Cys Lys Ser Arg Glu Asp Cys Ala Leu Phe
 60 500 505 510

Val Ser Glu Leu Tyr Gln Glu Leu Phe Ser Ala Ser Lys Tyr Ser Phe
515 520 525

5 Ile Asn Asp Asn Ala Ala Ser Gly Ile Xaa
530 535

10 (2) INFORMATION FOR SEQ ID NO: 329:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 202 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

Met Gly Ile Ala Leu Ala Val Leu Gly Trp Leu Ala Val Met Leu Cys
1 5 10 15

20 Cys Ala Leu Pro Met Trp Arg Val Thr Ala Phe Ile Gly Ser Asn Ile
20 25 30

25 Val Thr Ser Gln Thr Ile Trp Glu Gly Leu Trp Met Asn Cys Val Val
35 40 45

Gln Ser Thr Gly Gln Met Gln Cys Lys Val Tyr Asp Ser Leu Leu Ala
50 55 60

30 Leu Pro Gln Asp Leu Gln Ala Ala Arg Ala Leu Val Ile Ile Ser Ile
65 70 75 80

35 Ile Val Ala Ala Leu Gly Val Leu Leu Ser Val Val Gly Gly Lys Cys
85 90 95

Thr Asn Cys Leu Glu Asp Glu Ser Ala Lys Ala Lys Thr Met Ile Val
100 105 110

40 Ala Gly Val Val Phe Leu Leu Ala Gly Leu Met Val Ile Val Pro Val
115 120 125

Ser Trp Thr Ala His Asn Ile Ile Gln Asp Phe Tyr Asn Pro Leu Val
130 135 140

45 Ala Ser Gly Gln Lys Arg Glu Met Gly Ala Ser Leu Tyr Val Gly Trp
145 150 155 160

50 Ala Ala Ser Gly Leu Leu Leu Gly Gly Leu Leu Cys Cys Asn
165 170 175

Cys Pro Pro Arg Thr Asp Lys Pro Tyr Ser Ala Lys Tyr Ser Ala Ala
180 185 190

55 Arg Ser Ala Ala Ala Ser Asn Tyr Val Xaa
195 200

60 (2) INFORMATION FOR SEQ ID NO: 330:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 263 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

Met Ala Thr Val Thr Ala Thr Thr Lys Val Pro Glu Ile Arg Asp Val
1 5 10 15

10 Thr Arg Ile Glu Arg Ile Gly Ala His Ser His Ile Arg Gly Leu Gly
20 25 30

Leu Asp Asp Ala Leu Glu Pro Arg Gln Ala Ser Gln Gly Met Val Gly
35 40 45

15 Gln Leu Ala Ala Arg Arg Ala Ala Gly Val Val Leu Glu Met Ile Arg
50 55 60

20 Glu Gly Lys Ile Ala Gly Arg Ala Val Leu Ile Ala Gly Gln Pro Gly
65 70 75 80

Thr Gly Lys Thr Ala Ile Ala Met Gly Met Ala Gln Ala Leu Gly Pro
85 90 95

25 Asp Thr Pro Phe Thr Ala Ile Ala Gly Ser Glu Ile Phe Ser Leu Glu
100 105 110

Met Ser Lys Thr Glu Ala Leu Thr Gln Ala Phe Arg Arg Ser Ile Gly
115 120 125

30 Val Arg Ile Lys Glu Glu Thr Glu Ile Ile Glu Gly Glu Val Val Glu
130 135 140

Ile Gln Ile Asp Arg Pro Ala Thr Gly Thr Gly Ser Lys Val Gly Lys
35 145 150 155 160

Leu Thr Leu Lys Thr Thr Glu Met Glu Thr Ile Tyr Asp Leu Gly Thr
165 170 175

40 Lys Met Ile Xaa Ser Leu Thr Lys Asp Lys Val Gln Ala Gly Asp Val
180 185 190

Ile Thr Ile Asp Lys Ala Thr Gly Lys Ile Ser Lys Leu Gly Arg Ser
195 200 205

45 Phe Thr Arg Ala Arg Glu Leu Arg Arg Tyr Gly Leu Pro Asp Gln Val
210 215 220

Arg Ala Val Pro Arg Trp Gly Ala Pro Glu Thr Gln Gly Gly Ala
50 225 230 235 240

His Arg Val Pro Ala Arg Asp Arg Arg His Gln Leu Ser His Pro Gly
245 250 255

55 Leu Pro Gly Ala Leu Leu Arg
260

60 (2) INFORMATION FOR SEQ ID NO: 331:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 260 amino acids

(B) TYPE: amino acid

5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

Met Leu Ala Leu Leu Gly Leu Ser Gln Ala Leu Asn Ile Leu Leu Gly
 1 5 10 15

10 Leu Lys Gly Leu Ala Pro Ala Glu Ile Ser Ala Val Cys Glu Lys Gly
 20 25 30

15 Asn Phe Asn Val Ala His Gly Leu Ala Trp Ser Tyr Tyr Ile Gly Tyr
 35 40 45

Leu Arg Leu Ile Leu Pro Glu Leu Gln Ala Arg Ile Arg Thr Tyr Asn
 50 55 60

20 Gln His Tyr Asn Asn Leu Leu Arg Gly Ala Val Ser Gln Arg Leu Tyr
 65 70 75 80

25 Ile Leu Leu Pro Leu Asp Cys Gly Val Pro Asp Asn Leu Ser Met Ala
 85 90 95

Asp Pro Asn Ile Arg Phe Leu Asp Lys Leu Pro Gln Gln Thr Gly Asp
 100 105 110

30 Arg Ala Gly Ile Lys Asp Arg Val Tyr Ser Asn Ser Ile Tyr Glu Leu
 115 120 125

Leu Glu Asn Gly Gln Arg Ala Gly Thr Cys Val Leu Glu Tyr Ala Thr
 130 135 140

35 Pro Leu Gln Thr Leu Phe Ala Met Ser Gln Tyr Ser Gln Ala Gly Phe
 145 150 155 160

Ser Gly Glu Asp Arg Leu Glu Gln Ala Lys Leu Phe Cys Arg Thr Leu
 165 170 175

40 Glu Asp Ile Leu Ala Asp Ala Pro Glu Ser Gln Asn Asn Cys Arg Leu
 180 185 190

45 Ile Ala Tyr Gln Glu Pro Ala Asp Asp Ser Ser Phe Ser Leu Ser Gln
 195 200 205

Glu Val Leu Arg His Leu Arg Gln Glu Glu Lys Glu Glu Val Thr Val
 210 215 220

50 Gly Ser Leu Lys Thr Ser Ala Val Pro Ser Thr Ser Thr Met Ser Gln
 225 230 235 240

55 Glu Pro Glu Leu Leu Ile Ser Gly Met Glu Lys Pro Leu Pro Leu Arg
 245 250 255

Thr Asp Phe Ser
 260

(2) INFORMATION FOR SEQ ID NO: 332:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

Met Thr Pro Gln Lys Pro Ala Leu Ala Val Leu Leu Leu Glu Val Pro
10 1 5 10 15

Leu Leu Leu Thr Leu Ser Val Leu Lys Lys Arg Cys Leu Val Thr Cys
20 25 30

15 Glu Pro Thr Ser Arg Phe Val Ser Cys Asp Leu Pro Leu Ser Val Xaa
35 40 45

20

(2) INFORMATION FOR SEQ ID NO: 333:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

Met Ala Ala Ala Ala Trp Leu Gln Val Leu Pro Val Ile Leu Leu Leu
30 1 5 10 15

Leu Gly Ala His Pro Ser Pro Leu Ser Phe Phe Ser Ala Gly Pro Ala
35 20 25 30

Thr Val Ala Ala Ala Asp Arg Ser Lys Trp His Ile Pro Ile Pro Ser
35 40 45

40 Gly Lys Asn Tyr Phe Ser Phe Gly Lys Ile Leu Phe Arg Asn Thr Thr
50 55 60

Ile Phe Leu Lys Phe Asp Gly Glu Pro Cys Asp Leu Ser Leu Asn Ile
45 65 70 75 80

Thr Trp Tyr Leu Lys Ser Ala Asp Cys Tyr Asn Glu Ile Tyr Asn Phe
85 90 95

50 Lys Ala Glu Glu Val Glu Leu Tyr Leu Glu Lys Leu Lys Glu Lys Arg
100 105 110

Gly Leu Ser Gly Lys Tyr Gln Thr Ser Ser Lys Leu Phe Gln Asn Cys
115 120 125

55 Ser Glu Leu Phe Lys Thr Gln Thr Phe Ser Gly Asp Phe Met His Arg
130 135 140

Leu Pro Leu Leu Gly Glu Lys Gln Glu Ala Lys Glu Asn Gly Thr Asn
60 145 150 155 160

Leu Thr Phe Ile Gly Asp Lys Thr Ala Met His Glu Pro Leu Gln Thr
 165 170 175
 Trp Gln Asp Ala Pro Tyr Ile Phe Ile Val His Ile Gly Ile Ser Ser
 5 180 185 190
 Ser Lys Glu Ser Ser Lys Glu Asn Ser Leu Ser Asn Leu Phe Thr Met
 195 200 205
 10 Thr Val Glu Val Lys Gly Pro Tyr Glu Tyr Leu Thr Leu Glu Asp Tyr
 210 215 220
 Pro Leu Met Ile Phe Phe Met Val Met Cys Ile Val Tyr Val Leu Phe
 225 230 235 240
 15 Gly Val Leu Trp Leu Ala Trp Ser Ala Cys Tyr Trp Arg Asp Leu Leu
 245 250 255
 Arg Ile Gln Phe Trp Ile Gly Ala Val Ile Phe Leu Gly Met Leu Glu
 20 260 265 270
 Lys Ala Val Phe Tyr Ala Glu Phe Gln Asn Ile Arg Tyr Lys Gly Xaa
 275 280 285
 25 Ser Val Gln Gly Ala Leu Ile Leu Ala Glu Leu Leu Ser Ala Val Lys
 290 295 300
 Arg Ser Leu Ala Arg Thr Leu Val Ile Ile Val Ser Leu Gly Tyr Gly
 305 310 315 320
 30 Ile Val Lys Pro Arg Leu Glu Ser Leu Phe Ile Arg Leu Xaa
 325 330
 35

(2) INFORMATION FOR SEQ ID NO: 334:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 200 amino acids
 40 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:
 Met Val Leu Xaa Val Val Thr Leu Gly Leu Ala Leu Phe Thr Leu Cys
 45 1 5 10 15
 Gly Lys Phe Lys Arg Trp Lys Leu Asn Gly Ala Phe Leu Leu Ile Thr
 20 25 30
 50 Ala Phe Leu Ser Val Leu Ile Trp Val Ala Trp Met Thr Met Tyr Leu
 35 40 45
 Phe Gly Asn Val Lys Leu Gln Gln Gly Asp Ala Trp Asn Asp Pro Thr
 55 55 60
 55 Leu Ala Ile Thr Leu Ala Ala Ser Ala Gly Ser Ser Ser Ser Thr
 65 70 75 80
 60 Pro Ser Leu Arg Ser Thr Ala Pro Phe Cys Gln Pro Cys Arg Arg Thr
 85 90 95

Arg Pro Thr Thr Ser Thr Arg Arg Ser Pro Gly Cys Gly Arg Arg Pro
100 105 110

5 Ser Arg Arg Thr Cys Ser Cys Arg Gly Pro Ile Trp Arg Thr Arg Pro
115 120 125

Ser Pro Trp Met Asn Thr Met Gln Leu Ser Glu Gln Gln Asp Phe Pro
130 135 140

10 Thr Ala Ala Trp Glu Lys Asp Pro Val Ala Ala Trp Gly Lys Asp Pro
145 150 155 160

Ala Leu Arg Leu Glu Ala Thr Cys Ile Ser Gln Leu Arg Trp Pro Ser
15 165 170 175

Cys Ser Thr Val Gly Pro Ser Gln Leu Leu Arg Gln Val Thr Gln Glu
180 185 190

20 Xaa Thr Phe Gly Glu Arg Leu Xaa
195 200

25 (2) INFORMATION FOR SEQ ID NO: 335:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

Met Leu Leu His His Gln Leu Leu Ile Val Thr Leu His Leu Val Leu
1 5 10 15

35 Leu Leu Ala Thr Leu Leu Val Xaa
20

40 (2) INFORMATION FOR SEQ ID NO: 336:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

50 Met Thr Lys Ala Leu Leu Ile Tyr Leu Val Ser Ser Phe Leu Ala Leu
1 5 10 15

Asn Gln Ala Ser Leu Ile Ser Arg Cys Asp Leu Ala Gln Val Leu Gln
20 25 30

55 Leu Glu Asp Leu Asp Gly Phe Glu Gly Tyr Ser Leu Ser Asp Trp Leu
35 40 45

Cys Leu Ala Phe Val Glu Ser Lys Phe Asn Ile Ser Lys Ile Asn Glu
50 55 60

Asn Ala Asp Gly Ser Phe Asp Tyr Gly Leu Phe Gln Ile Asn Ser His
 65 70 75 80

5 Tyr Trp Cys Asn Xaa Tyr Lys Ser Tyr Ser Glu Asn Leu Cys His Val
 85 90 95

Asp Cys Gln Asp Leu Leu Asn Pro Asn Leu Leu Ala Gly Ile His Cys
 100 105 110

10 Ala Lys Arg Ile Val Ser Gly Ala Arg Gly Met Asn Asn Trp Val Arg
 115 120 125

Met Glu Xaa Cys Thr Val Gln Ala Gly His Ser Ser Thr Gly Xaa
 130 135 140

15

(2) INFORMATION FOR SEQ ID NO: 337:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

25

Met Leu Val Ile Ala Gly Gly Ile Leu Ala Ala Leu Leu Leu Ile
 1 5 10 15

30

Val Val Val Leu Cys Leu Tyr Phe Lys Ile His Asn Ala Leu Lys Ala
 20 25 30

Ala Lys Glu Pro Glu Ala Val Ala Val Lys Asn His Asn Pro Asp Lys
 35 40 45

35

Val Trp Trp Ala Lys Asn Ser Gln Ala Lys Thr Ile Ala Thr Glu Ser
 50 55 60

40

Cys Pro Ala Leu Gln Cys Cys Glu Gly Tyr Arg Met Cys Ala Ser Phe
 65 70 75 80

Asp Ser Leu Pro Pro Cys Cys Cys Asp Ile Asn Glu Gly Leu Xaa
 85 90 95

45

(2) INFORMATION FOR SEQ ID NO: 338:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

50

Met Leu Leu Lys Ser Asn Ile Leu Met Leu Asn Leu Phe Ala Ala Asn
 55 1 5 10 15

60

Val Gly Ala Asn Phe Ala Leu Thr Val Glu Lys Ile Gly Met Ile Leu
 20 25 30

Leu Asn Val Ser Gly Xaa

35

5 (2) INFORMATION FOR SEQ ID NO: 339:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

Met Leu Val Val Ala Phe Gly Leu Leu Val Leu Tyr Ile Leu Leu Ala
1 5 10 15

Ser Ser Trp Lys Arg Pro Glu Pro Gly Ile Leu Thr Asp Arg Gln Pro
20 25 30

Leu Leu His Asp Gly Glu Xaa
20 35

25 (2) INFORMATION FOR SEQ ID NO: 340:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

Ser Asp Pro Leu Ala Ser Ala Ser Gln Asn Ala Gly Ile Val Ser Val
1 5 10 15

Gly Leu Cys Thr Arg Pro Gly Pro Gln Phe Lys Asn Ala Gln Pro Pro
20 25 30

Phe Pro Xaa Gln Lys Ala Pro Arg Cys Leu Trp Glu Asn Gln Pro Pro
35 40 45

Pro Trp Arg Lys Ala Trp Asp Leu Pro Ser His Leu Gly Arg Arg Gly
50 55 60

Ile Cys Gly Lys Ser Phe Xaa
65 70

50 (2) INFORMATION FOR SEQ ID NO: 341:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 85 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

Tyr Val Met Ile Phe Lys Lys Glu Phe Ala Pro Ser Asp Glu Glu Leu
1 5 10 15

Asp Ser Tyr Arg Arg Gly Glu Glu Trp Asp Pro Gln Lys Ala Glu Glu

	20	25	30	
	Lys Arg Asn Xaa Lys Glu Leu Ala Gln Arg Gln Xaa Gly Gly Gly Ser			
	35	40	45	
5	Pro Ala Gly Ala Cys Gly Gly Glu Pro Cys Gln Arg Leu Gln Gly Gln			
	50	55	60	
	Val Gln Pro Pro His Arg Gln Gly Ser Ser Gln Arg Arg Ser Pro His			
10	65	70	75	80
	Ala Thr Gly Gln Xaa			
	85			

15

(2) INFORMATION FOR SEQ ID NO: 342:

	(i) SEQUENCE CHARACTERISTICS:			
20	(A) LENGTH: 90 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:			
25	Met Trp Asp Trp Asp Trp Ser Ala Pro Trp Ser Trp Pro Leu Trp Leu			
	1	5	10	15
	Ser Leu Ala Leu Val Cys Leu Ser Ala Gly Ala Lys Gly His Arg Ala			
30	20	25	30	
	Ser Glu Ala Gly His Ala Arg Ala Leu Thr Cys Glu Met Gly Ser Glu			
	35	40	45	
35	Phe Xaa Thr Ala Xaa Gly Leu Val Leu Gly Xaa Xaa Xaa Trp Thr Xaa			
	50	55	60	
	Xaa Asn Gly Ser Ala Gly Pro Glu Arg Arg Gly Trp Arg Pro Ala Ala			
	65	70	75	80
40	Phe Leu Ala Val Phe Leu Leu Gly Asp Xaa			
	85	90		

45 (2) INFORMATION FOR SEQ ID NO: 343:

	(i) SEQUENCE CHARACTERISTICS:			
50	(A) LENGTH: 48 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:			
55	Met Phe Gly Pro Thr Phe His Ser Leu Val Leu Val Pro Pro Trp Pro			
	1	5	10	15
	Asn Leu Ser Leu Leu His Phe Thr Ser Pro Val Gly Gln His Ser Ser			
	20	25	30	
60	Phe Leu Pro Thr Ser Leu Arg Leu Xaa Lys Lys Lys Lys Lys Lys			
	35	40	45	

5

(2) INFORMATION FOR SEQ ID NO: 344:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

15 Met Cys Ser Lys Asn Gly Phe Leu Leu Ala Trp Ser Trp Asn Ser Pro
 1 5 10 15

 Trp Leu Pro Gln Ala Ser Leu Ala His Gly Cys Trp Gly Arg Trp Met
 20 25 30

 20 Ser Asp Leu Val Gly Cys Ser Arg Glu Asn Lys Cys Ala Leu Arg Asp
 35 40 45

 His Ser Glu Arg Val Gln Gly Xaa
 50 55

(2) INFORMATION FOR SEO ID NO: 345:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 222 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

35 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 345.

Lys Met Leu Lys Asn Leu Arg Asn Thr Asn Tyr Asp Val Phe Glu Lys
 130 135 140

Ile Cys Trp Gly Leu Gly Ile Glu Tyr Thr Phe Pro Pro Leu Tyr Tyr
 5 145 150 155 160

Arg Arg Ala His Arg Arg Phe Val Thr Lys Lys Ala Leu Cys Ile Arg
 165 170 175

10 Val Phe Gln Glu Thr Gln Lys Leu Lys Lys Arg Arg Ala Leu Lys
 180 185 190

Ala Ala Ala Ala Gln Lys Gln Ala Lys Arg Arg Asn Pro Asp Ser
 15 195 200 205

Pro Ala Lys Ala Ile Pro Lys Thr Leu Lys Asp Ser Gln Xaa
 210 215 220

20 (2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:
 25 (A) LENGTH: 64 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

30 Met Gly Ala Pro Ala Ala Ser Leu Leu Leu Leu Leu Leu Phe Ala
 1 5 10 15

Cys Cys Trp Ala Pro Gly Gly Ala Asn Leu Ser Gln Asp Asp Ser Gln
 35 20 25 30

Pro Trp Thr Ser Asp Glu Thr Val Val Ala Gly Gly Thr Val Val Leu
 35 40 45

Lys Cys Gln Val Lys Asp His Glu Asp Ser Ser Leu Gln Trp Ser Xaa
 40 50 55 60

45 (2) INFORMATION FOR SEQ ID NO: 347:

(i) SEQUENCE CHARACTERISTICS:
 50 (A) LENGTH: 154 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

55 Met Val Ala Pro Val Trp Tyr Leu Val Ala Ala Ala Leu Leu Val Gly
 1 5 10 15

Phe Ile Leu Phe Leu Thr Arg Ser Arg Gly Arg Ala Ala Ser Ala Gly
 60 20 25 30

Gln Glu Pro Leu His Asn Glu Glu Leu Ala Gly Ala Gly Arg Val Ala

35 40 45

Gln Pro Gly Pro Leu Glu Pro Glu Glu Pro Arg Ala Gly Gly Arg Pro
 50 55 60

5 Arg Arg Arg Arg Asp Leu Gly Ser Arg Leu Gln Ala Gln Arg Arg Ala
 65 70 75 80

10 Gln Arg Val Ala Trp Ala Glu Ala Asp Glu Asn Glu Glu Ala Val
 10 85 90 95

Ile Leu Ala Gln Glu Glu Gly Val Glu Lys Pro Ala Glu Xaa His
 100 105 110

15 Leu Ser Gly Lys Ile Gly Ala Lys Lys Leu Arg Xaa Xaa Glu Glu Lys
 115 120 125

Gln Ala Arg Lys Ala Gln Xaa Glu Ala Glu Glu Ala Glu Arg Glu Xaa
 130 135 140

20 Arg Lys Arg Leu Glu Ser Gln Arg Glu Xaa
 145 150

25 (2) INFORMATION FOR SEQ ID NO: 348:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

35 Met Gln Lys Cys Met Leu Ser Ala Leu Val Phe His Ile Gln Trp Ser
 1 5 10 15

Xaa

40

(2) INFORMATION FOR SEQ ID NO: 349:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

50 Met Leu Val Cys Ser Phe Leu Phe Leu Xaa
 1 5 10

55 (2) INFORMATION FOR SEQ ID NO: 350:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Val Ile Glu Leu Cys Val Ser Leu Arg Ser Leu Asn Phe Xaa
1 5 10

5

(2) INFORMATION FOR SEQ ID NO: 351:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

15

Met Cys Glu Phe Xaa Xaa Xaa Ile Met Xaa Leu Ala Gly Tyr Phe Ala
1 5 10 15

20

Cys Xaa

(2) INFORMATION FOR SEQ ID NO: 352:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

Met Val Gly Gly Tyr Val Ser Ser Phe Ser Phe Pro Pro Val Ser Ser
1 5 10 15

35

Ser Leu Leu Leu Pro Ala Ser Phe Ala Phe Pro Phe Leu Pro Gly Thr
20 25 30

Pro Cys Pro Phe Leu Tyr Phe Leu Pro Ser Pro Phe Ser Pro Leu Pro
35 40 45

40

Leu Ser Leu Thr Arg Ser Asn Ser Phe Leu Leu Asn Gly Xaa
50 55 60

45

(2) INFORMATION FOR SEQ ID NO: 353:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

50

Glu Lys Lys Ser Met Ser Val Ser Asp Ile Tyr Ala Leu Glu Ser Leu
1 5 10 15

Gly Arg Ser Leu Phe Thr Leu Asn Ser Met Cys Leu Pro Leu Ser Phe
20 25 30

60

Xaa

5 (2) INFORMATION FOR SEQ ID NO: 354:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 245 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

Met Gly Gly Ala Ser Arg Arg Val Glu Ser Gly Ala Trp Ala Tyr Leu
1 5 10 15

Ser Pro Leu Val Leu Arg Lys Glu Leu Glu Ser Leu Val Glu Asn Glu
20 25 30

Gly Ser Glu Val Leu Ala Leu Pro Glu Leu Pro Ser Ala His Pro Ile
20 35 40 45

Ile Phe Trp Asn Leu Leu Trp Tyr Phe Gln Arg Leu Arg Leu Pro Ser
50 55 60

25 Ile Leu Pro Gly Leu Val Leu Ala Ser Cys Asp Gly Pro Ser Xaa Ser
65 70 75 80

Gln Ala Pro Ser Pro Trp Leu Thr Pro Asp Pro Ala Ser Val Gln Val
85 90 95

30 Arg Leu Leu Trp Asp Val Leu Thr Pro Asp Pro Asn Ser Cys Pro Pro
100 105 110

Leu Tyr Val Leu Trp Arg Val His Ser Gln Ile Pro Gln Arg Val Val
35 115 120 125

Trp Pro Gly Pro Val Pro Ala Ser Leu Ser Leu Ala Leu Leu Glu Ser
130 135 140

40 Val Leu Arg His Val Gly Leu Asn Glu Val His Lys Ala Val Gly Leu
145 150 155 160

Leu Leu Glu Thr Leu Gly Pro Pro Pro Thr Gly Leu His Leu Gln Arg
165 170 175

45 Gly Ile Tyr Arg Glu Ile Leu Phe Leu Thr Met Ala Ala Leu Gly Lys
180 185 190

50 Asp His Val Asp Ile Val Ala Phe Asp Lys Lys Tyr Lys Ser Ala Phe
195 200 205

Asn Lys Leu Ala Ser Ser Met Gly Lys Glu Glu Leu Arg His Arg Arg
210 215 220

55 Ala Gln Met Pro Thr Pro Lys Ala Ile Asp Cys Arg Lys Cys Phe Gly
225 230 235 240

Ala Pro Pro Glu Cys
245

(2) INFORMATION FOR SEQ ID NO: 355:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

10 Met Lys Phe Ser Leu Leu Phe Leu Pro Met Leu Leu Ile Leu Lys Pro
1 5 10 15

15 Asp Leu Phe His Ile Ser Ile Cys Thr Leu Ala Ala Cys Gly Leu Thr
20 25 30

Phe Pro Xaa
35

20

(2) INFORMATION FOR SEQ ID NO: 356:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

30 Met Leu Phe Phe Phe Ile Leu His Leu Leu Ser Ile Met Ser Phe Leu
1 5 10 15

Ser Pro Asp Ile Met Xaa
20

35

(2) INFORMATION FOR SEQ ID NO: 357:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 98 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

45 Met Phe Gly Leu Leu Val Glu Ser Gln Thr Leu Leu Glu Glu Asn Ala
1 5 10 15

50 Val Gln Gly Thr Glu Arg Thr Leu Gly Leu Asn Ile Ala Pro Phe Ile
20 25 30

Asn Gln Phe Gln Val Pro Ile Arg Val Phe Leu Asp Leu Ser Ser Leu
35 40 45

55 Pro Cys Ile Pro Leu Ser Lys Pro Val Glu Leu Leu Arg Leu Asp Leu
50 55 60

60 Met Thr Pro Tyr Leu Asn Thr Ser Asn Arg Glu Val Lys Val Tyr Val
65 70 75 80

Cys Xaa Ile Trp Glu Asp Leu Thr Ala Ile Pro Phe Trp Val Ser Tyr
85 90 95

Val Pro

5

10 (2) INFORMATION FOR SEQ ID NO: 358:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

Met Phe Gly Ala His Arg Xaa Trp Gln Gly Ser Val Leu Leu Phe Leu
1 5 10 15

20 Ser Phe Ala Trp Gly Asn Gly Gly Ser Val Thr Phe Ser Asp Val Pro
20 25 30

Arg Val Met Pro Leu Ala Gly Gly Pro Xaa Xaa Gln Val Ser Ser Thr
35 40 45

25 Pro Arg Pro Pro Pro His Gln Val Thr Ser Ser Pro Gly Leu Glu Ser
50 55 60

30 Ala His Ile Val Cys Pro Glu Arg Lys Lys Lys Lys Lys
65 70 75

35 (2) INFORMATION FOR SEQ ID NO: 359:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

Thr Leu Leu Xaa Phe Leu Xaa Leu Leu Thr Thr Glu Gly Gly Arg Glu
1 5 10 15

45 Asn Ile Phe Xaa Gly Arg Ile Leu Xaa Leu Gln Xaa Ser Pro Xaa
20 25 30

50 (2) INFORMATION FOR SEQ ID NO: 360:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

Met Leu Ser Phe Phe Ile Cys Leu Leu Ile Phe Val His Leu Leu
1 5 10 15

60